CÁLCULO DIFERENCIAL E INTEGRAL EN VARIAS VARIABLES

Notas de curso

Marcelo Fiori

Febrero 2019

Introducción

Estas notas fueron preparadas entre 2017 y 2018, para servir como material de estudio en el curso de Cálculo Diferencial e Integral en Varias Variables, de la Facultad de Ingeniería, UdelaR.

A lo largo de estas notas hay contenido encerrado en cuadros de colores. Esto responde a una opinión personal del autor sobre la importancia de algunos conceptos en el marco del curso de Cálculo Diferencial e Integral en Varias Variables.

Importante

El contenido en estas cajas es sumamente importante, y muchas veces presenta conceptos fundamentales.
No se puede pasar por este curso sin comprender estas ideas.

Extra

El contenido en estas cajas puede ser anecdótico, o pueden ser comentarios más profundos que lo esperado.
Se puede pasar por este curso sin estudiar a fondo estas ideas.

Un texto de matemática no está pensado para leerse a la misma velocidad que una novela o un libro de cuentos. Lea las definiciones y los comentarios con atención y tranquilidad. Lea las demostraciones despacio. Luego intente resumir cuáles son las ideas principales, intente volver a recorrer el camino lógico en su cabeza, y si es necesario vuelva a leer la prueba.

Hay dos objetivos fundamentales en los cursos de matemática de una Facultad de Ingeniería. El primero consiste en que el estudiante maneje las herramientas matemáticas contenidas en el programa. Por ejemplo, al finalizar el curso el estudiante debe ser capaz de calcular integrales o desarrollos de Taylor.

El segundo, y probablemente más importante, es la "madurez matemática". La capacidad del estudiante para enfrentarse a problemas y conceptos nuevos, que es fundamental en carreras de ciencia como lo son las ingenierías.

No importa si dentro de dos años no nos acordamos de la demostración del Teorema de Fubini. Lo importante es haber sido capaces de comprenderlo en una oportunidad (y eventualmente volver a hacerlo si fuera necesario).

Esta es la primera versión de este material, que deberá pasar necesariamente por un proceso de revisión y re-edición. Por lo tanto, comentarios, errores, y sugerencias sobre estas notas son más que bienvenidos. Están invitados a compartirlos a mfiori@fing.edu.uy.

Marcelo Fiori
Febrero de 2019

Índice general

1. Números complejos **5**

1.1. Introducción y propiedades básicas 5

1.2. Exponencial compleja 13

1.3. Raíces complejas 18

2. Sucesiones y Series **21**

2.1. Sucesiones 21

2.1.1. Sucesiones y completitud 31

2.1.2. Sucesiones y continuidad 32

2.2. Series 34

2.2.1. Series de términos positivos 38

2.2.2. Series alternadas 42

3. Integrales impropias **45**

3.1. Integrales impropias de primera especie 45

3.2. Integrales impropias de segunda especie 54

3.3. Integrales mixtas 56

4. Topología y sucesiones en $\mathbb{R}^n$ **59**

4.1. Topología de $\mathbb{R}^n$ 59

4.2. Sucesiones en R^n 69

5. Continuidad **75**

5.1. Funciones en $\mathbb{R}^n$ 75

5.2. Límites y continuidad 79

5.3. Teoremas para funciones continuas 88

6. Diferenciabilidad **91**

6.1. Derivadas parciales y direccionales 91

6.2. Diferenciabilidad 97

6.3. Derivadas de orden superior 109

6.4. Funciones a $\mathbb{R}^m$ 112

6.5. Desarrollo de Taylor 117

6.6. Extremos relativos 122

7. Integrales múltiples **127**

7.1. Cambios de Variable 140

7.1.1. Coordenadas Polares 146

7.1.2. Coordenadas Cilíndricas 152

7.1.3. Coordenadas Esféricas 154

Capítulo 1

Números complejos

Comenzaremos este texto trabajando con los números complejos, que si bien serán de utilidad en algunos de los próximos capítulos, es el capítulo que tiene menos contenido de cálculo en sí, ya que no hay límites, continuidad, ni derivadas.

1.1. Introducción y propiedades básicas

De los conjuntos numéricos que estudiaremos en este curso ($\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$), el de los números complejos es quizás el único que tiene un origen histórico bien marcado, y que además es muy interesante.

El comienzo se da en Italia en el siglo *XVI*. Primero Gerolamo Cardano publica un libro describiendo una solución algebraica para hallar raíces de polinomios de tercer y cuarto grado[1]. Cuando se aplicaba este método al ejemplo histórico $x^3 = 15x + 4$, la expresión resultaba $x = \sqrt[3]{2+\sqrt{-121}} + \sqrt[3]{2-\sqrt{-121}}$. En este punto, Cardano decía que la fórmula no era aplicable en estos casos, pues aparecía $\sqrt{-121}$ que no tenía sentido. Fue el también italiano Rafael Bombelli quien comenzó a manipular algebraicamente estas expresiones, pasando por alto que no tenían sentido, y llegó a que

$$\sqrt[3]{2+\sqrt{-121}} + \sqrt[3]{2-\sqrt{-121}} = 4,$$

que era efectivamente una solución conocida de $x^3 = 15x + 4$.

Bombelli listó las operaciones aritméticas con estas expresiones, que serían las reglas para operar con estos números "imaginarios" como los llamó René Descartes, debido a que "su naturaleza es imposible [...] y solamente existen en la imaginación", como decía Euler.

[1] Aunque la fórmula en realidad fue comunicada a Cardano por otros matemáticos italianos. Ver nota histórica al final de este capítulo

Muchas veces se pone a la ecuación cuadrática $x^2+1=0$ como el origen histórico de los números complejos. Sin embargo, fueron las ecuaciones cúbicas las que llevaron al desarrollo de estos números.

Se trabajaba entonces con expresiones como $2+3i$ (aunque la notación no era esta), y se operaba con ellas como si fueran números reales, con la particularidad que $i^2=-1$. Así, el producto de dos complejos sería:

$$(a+bi)(c+di)=ac+adi+bci+bdi^2=(ac-bd)+(ad+bc)i. \tag{1.1}$$

Desde el punto de vista operativo, esto es suficiente para poder trabajar con los números complejos. De hecho, durante casi tres siglos, muchos matemáticos (incluyendo a Leibnitz, Bernoulli, D'Alembert, Lagrange, y Euler, quien fue el que introdujo la notación i para la unidad imaginaria) utilizaron estos números, manipulando las expresiones según las reglas dadas pero sin una definición formal, dando lugar a resultados de gran importancia. Fueron Gauss y Hamilton los que dieron una definición algebraica rigurosa de los números complejos como pares ordenados de reales, que es la que se utiliza hasta hoy.

Daremos entonces esta definición formal, pero rápidamente volveremos a utilizar la notación más cómoda de $a+bi$, sabiendo que todas las operaciones y propiedades están rigurosamente fundamentadas.

Definición 1.1. Un número complejo es un par ordenado de números reales (a,b).

Observar que, al ser un par ordenado, no es lo mismo el número $(2,3)$ que $(3,2)$.

Dado un número complejo $z=(a,b)$, su primera componente a se denomina parte real, y se denota $a=Re(z)$. La segunda componente se denomina parte imaginaria, y se escribe $b=Im(z)$.

Las operaciones se definen para que este conjunto copie el comportamiento que esperamos (es decir, para que el producto sea en definitiva el de la ecuación (1.1)).

Definición 1.2. Sean (a,b) y (c,d) dos números complejos. Entonces:

- Decimos que $(a,b)=(c,d)$ sii $a=c$ y $b=d$.
- Definimos la suma como $(a,b)+(c,d)=(a+c,b+d)$.
- Definimos el producto como $(a,b)\cdot(c,d)=(ac-bd,ad+bc)$.

Es fácil chequear que las operaciones así definidas son asociativas, conmutativas, y cumplen la propiedad distributiva. También se puede verificar que el elemento $(0,0)$ es el neutro de la suma, que $(1,0)$ es el neutro del producto, y que tenemos opuesto e inverso. Es decir, el conjunto de números complejos, con las operaciones que recién definimos, cumple con todos los Axiomas de cuerpo.

Ejercicio 1.3
Verificar que, dado un complejo (a,b), su opuesto está dado por $(-a,-b)$, y en caso de ser $(a,b) \neq (0,0)$, su inverso está dado por $\left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right)$.

Notar que en la definición no aparece el símbolo i ni $\sqrt{-1}$. Es decir, hemos hecho una definición algebraica, formal, que describe el *comportamiento* de los números complejos, las reglas de manipulación que se usaron desde el siglo *XVI*, pero sin incluir explícitamente las raíces de números negativos.

Observemos también que, al estar definidos como pares ordenados de números reales, no podemos ver directamente la inclusión de los números reales en este nuevo conjunto, como sí lo podíamos hacer con $\mathbb{N}$ en $\mathbb{Q}$, o con $\mathbb{Q}$ en $\mathbb{R}$ por ejemplo. Sin embargo, podemos *identificar* a los reales con los números complejos de la forma $(a,0)$. En efecto, si tomamos dos números de esa forma, $(a,0)$ y $(b,0)$, su suma $(a+b,0)$ y su producto $(ab,0)$ también tienen componente imaginaria nula, y copia la estructura operativa que teníamos en los reales. Si llamamos C_0 a este conjunto de números complejos que solamente tienen parte real, entonces tenemos una función $f : \mathbb{R} \to C_0$ que justamente a cada real a le asigna el número complejo $(a,0)$. Tenemos entonces una copia[2] de los números reales adentro de $\mathbb{C}$, y en este sentido entonces, podemos pensar que estamos agrandando el cuerpo de los números reales. A partir de ahora, cuando un número complejo tenga parte imaginaria nula (es decir, es un número de esta copia de los reales que llamamos C_0), vamos a escribirlo simplemente con su parte real: $(a,0) = a$, pero debemos tener en cuenta que estamos haciendo un abuso de notación. Escribiremos por ejemplo $z = 2$, pero sin olvidar que es en realidad el complejo $(2,0)$.

Ahora sí, llamemos i al número complejo $(0,1)$. Observemos que si multiplicamos al complejo $(0,1)$ por sí mismo, resulta

$$(0,1)\cdot(0,1) = (0\cdot 0 - 1\cdot 1, 0\cdot 1 + 1\cdot 0) = (-1,0) = -1.$$

Es decir, $i^2 = -1$, y por lo tanto i es solución de $x^2+1=0$. Esta ecuación, como sabemos, no tiene solución en los reales. La situación es más extrema todavía: todo polinomio de grado n en los complejos, tiene n raíces. A este resultado se lo conoce como Teorema Fundamental del Álgebra, y la primeras demostraciones esencialmente correctas las dieron D'Alembert (1746) y Gauss (1799), pero no lo demostraremos en este curso. Se dice entonces que el conjunto de los números complejos es un cuerpo algebraicamente cerrado[3].

[2] Esta función f conserva la estructura: lleva, por ejemplo, el producto en $\mathbb{R}$ en el producto de las imágenes en $\mathbb{C}$. Es en realidad lo que se denomina un isomorfismo, tanto $\mathbb{R}$ como C_0 tienen la misma estructura algebraica. Son copias idénticas, que se comportan igual, pero donde cada número tiene un "nombre" distinto en cada conjunto: a en uno, $(a,0)$ en el otro.

[3] Es decir, ya no hay que seguir agrandando el conjunto para poder resolver ecuaciones polinomiales. Cuando trabajamos en $\mathbb{Q}$, la ecuación $x^2-2=0$ no tiene solución, entonces agrandamos el conjunto a $\mathbb{R}$. Pero ahora la ecuación $x^2+1=0$ no tiene solución, entonces agrandamos el conjunto a $\mathbb{C}$, y ahora todas las ecuaciones tienen solución en los complejos.

Tenemos entonces un cuerpo $\mathbb{C}$, que es una extensión de los reales, pero que se porta mejor con las ecuaciones polinomiales. ¿Perdimos algo en este camino? Es decir, ¿hay alguna propiedad de los reales que hayamos perdido en los complejos? Sí, lo que no podemos hacer en $\mathbb{C}$ es ordenarlos, de manera que el orden se "porte bien" con las operaciones. Los reales son entonces un cuerpo ordenado, y los complejos no. Pero los complejos son un cuerpo algebraicamente cerrado, y los reales no.

Con ese abuso de notación que nos permite escribir, por ejemplo $(2,0) = 2$, podemos escribir un complejo cualquiera $z = (a,b)$ descomponiendo en su parte real y su parte imaginaria:

$$z = (a,b) = (a,0) + (0,b) = a(1,0) + b(0,1) = a + bi.$$

A esta forma de escribir números complejos se le llama notación binómica.

La definición como par ordenado ya da una idea geométrica clara: interpretamos un número complejo como un punto en el plano, con coordenadas (a,b). Es decir, en el eje horizontal representamos la parte real, y en el eje vertical la parte imaginaria.

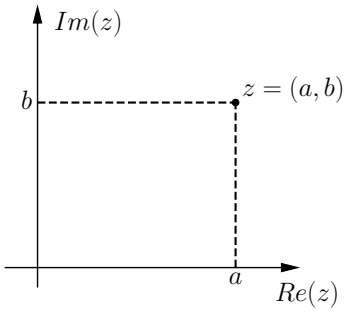


Figura 1.1: Representación del complejo $z = (a,b)$ en el plano.

Con esta interpretación, la suma de complejos no es más que la suma de vectores en el plano (o la regla del paralelogramo).

Identificamos a los complejos entonces como puntos en el plano, y los representamos mediante sus coordenadas, que son las partes real e imaginaria del complejo. Pero hay otras maneras de describir los puntos del plano, por ejemplo mediante su distancia al origen, y el ángulo que forman con el eje horizontal. Esto lleva a la siguiente definición y representación de los complejos:

Definición 1.4. Sea $z = a + bi$ un complejo. Llamamos *módulo* a la distancia de z al origen: $\sqrt{a^2+b^2}$, y se denota $|z|$. Al ángulo φ que forma el vector (a,b) con el eje horizontal lo denominamos *argumento*.

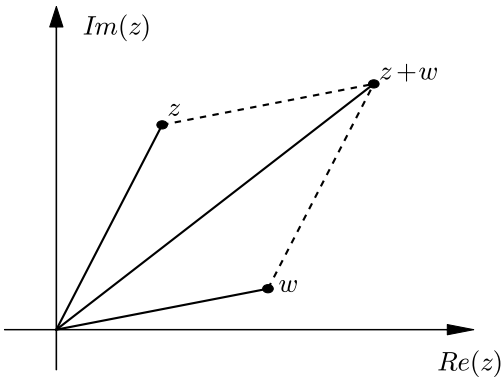


Figura 1.2: Suma de complejos.

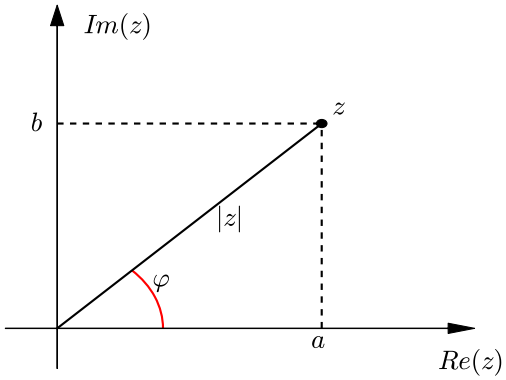


Figura 1.3: Módulo $|z|$ y argumento φ de un complejo z.

Observar que hay varios argumentos posibles, dado que el ángulo está determinado a menos de múltiplos de 2π. Es decir, podemos dar vueltas enteras y llegar al mismo punto en el plano: el complejo que tiene módulo 1 y argumento $\pi/20$ es el mismo que el complejo de módulo 1 y argumento $\pi/20+4\pi$. Para calcular el argumento de un complejo a partir de su expresión en notación binómica, notar que la tangente del ángulo que forma z con el eje real es $\frac{b}{a}$, es decir, $\varphi = \arctan\left(\frac{b}{a}\right)$. Pero hay que tener cuidado con esto:

Ejemplo 1.5
El complejo $z_1 = 1+i$ *tiene módulo* $|z_1| = \sqrt{1^2+1^2} = \sqrt{2}$*, y argumento* $\varphi_1 = \arctan\left(\frac{1}{1}\right) = \arctan(1) = \frac{\pi}{4}$*. Esto ya lo intuíamos a partir de la representación gráfica.*

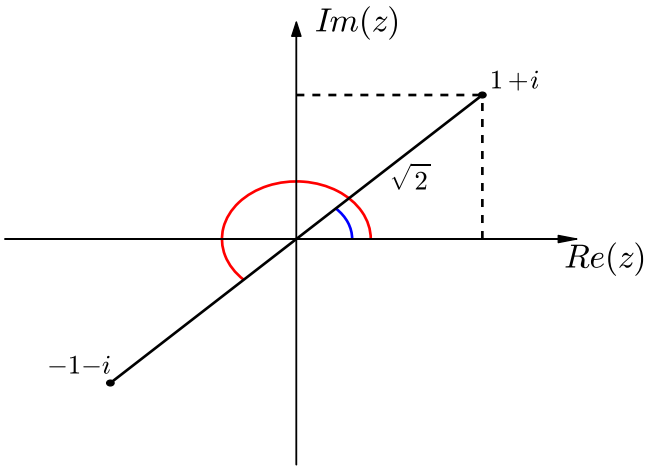


Figura 1.4: Complejos $1+i$ y $-1-i$.

Tomemos ahora el complejo $z_2 = -1 - i$.
Entonces su módulo es $|z_2| = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2}$. Cuando intentamos calcular su argumento, si no tenemos cuidado, resulta: $\varphi_2 = \arctan\left(\frac{-1}{-1}\right) = \arctan(1) = \frac{\pi}{4}$. Es decir, tendría el mismo módulo y argumento que z_1, por lo que algo debe estar mal, pues dos complejos con el mismo módulo y mismo argumento, necesariamente son iguales. Además, el resultado no tiene sentido con la representación gráfica que teníamos.

El problema es que la tangente tiene varias ramas, y estamos usando (en ambos casos) la que está definida en el intervalo $(-\frac{\pi}{2}, \frac{\pi}{2})$, por lo que en algunos casos hay que corregir, sumando (o restando) π. En el caso de z_2 por ejemplo, debíamos sumar π al resultado. En general lo que se debe hacer es corroborar si el resultado tiene sentido con la representación gráfica, mirando en qué cuadrante se encuentra el complejo.

El módulo de un complejo es la distancia al origen, y por lo tanto no puede ser un valor negativo, por ejemplo. Listemos algunas de estas propiedades, que serán estudiadas más en detalle, y en un marco más general, en el capítulo 4.

Proposición 1.6. Sean z y w dos complejos. Entonces:

1. $|z| \geqslant 0$, y $|z| = 0$ sii $z = 0$
2. $|Re(z)| \leqslant |z|$; $|Im(z)| \leqslant |z|$.
3. $|zw| = |z||w|$; $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$.
4. $|z+w| \leqslant |z| + |w|$.

Ejercicio 1.7
Interpretar geométricamente las propiedades 2 y 4.

Ejercicio 1.8
Observar que, como el módulo de un complejo mide la distancia al origen, entonces $|z - z_0|$ es la distancia de z al complejo z_0. Si consideramos entonces los complejos z que verifican $|z - i| = 1$ (es decir, que distan exactamente 1 del complejo i), ¿qué forma geométrica determinan?. Escribir z en su notación binómica $z = a + bi$, y reconocer la ecuación resultante.

Si tenemos un número complejo dado con su módulo $|z|$ y ángulo φ, podemos calcular su parte real e imaginaria, simplemente proyectando el vector a cada uno de los ejes (ver de nuevo la Figura 1.3):

$$a = Re(z) = |z| \cos\varphi, \qquad b = Im(z) = |z| \sin\varphi. \tag{1.2}$$

Ejercitemos estos cálculos con algunos ejemplos.

Ejemplos 1.9

1. *El complejo $z_1 = 1$ tiene módulo 1 y argumento 0 (o 2π, o cualquier múltiplo de 2π).*
2. *El complejo $z_2 = -1$ tiene módulo 1 y argumento π (o $-\pi$, etc). En general, los números reales (con parte imaginaria nula), tienen argumento 0 si son positivos, o π si son negativos.*
3. *El complejo $z_3 = i$, la unidad imaginaria, tiene módulo $|i| = \sqrt{0^2 + 1^2} = 1$, y argumento $\pi/2$.*
4. *El complejo $z_4 = -i$ tiene módulo 1, y argumento $-\pi/2$ (o $3\pi/2$).*
5. *El complejo $z_5 = 1 - i$ tiene módulo $|1 - i| = \sqrt{1^2 + 1^2} = \sqrt{2}$, y argumento $-\pi/4$.*

Observar en particular que los números complejos 1, -1, i y $-i$ están ubicados en los vértices de un cuadrado (¡dibújelos!). Cuando un número complejo tiene parte real nula, como por ejemplo $2i$, se dice que es imaginario puro.

Ejercicio 1.10
¿Qué ángulos tienen los números reales como argumento? ¿Y los imaginarios puros?

Al poder representar los complejos en el plano, hay muchas propiedades geométricas que aparecen naturalmente. Hay una transformación en particular, que es la de simetrizar respecto al eje real, que tiene un nombre y notación concretas:

Definición 1.11. Dado un complejo $z = a + bi$, definimos su conjugado como $\bar{z} = a - bi$.

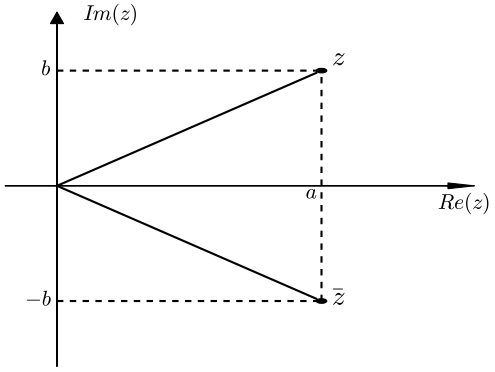


Figura 1.5: Conjugado $\bar{z}$ de un complejo.

Es una operación sumamente sencilla, y las demostraciones de todas las propiedades que listamos a continuación son simples cálculos, que quedan de ejercicio.

Proposición 1.12. Sean z y w dos complejos. Entonces:

1. $\overline{z+w} = \overline{z} + \overline{w}$
2. $\overline{zw} = \overline{z} \cdot \overline{w}$
3. $z + \overline{z} = 2Re(z)$
4. $z - \overline{z} = 2i \cdot Im(z)$
5. $\overline{\overline{z}} = z$
6. $z \cdot \overline{z} = |z|^2$

La última propiedad en particular, permite realizar cálculos de cocientes de complejos de forma sencilla. Si queremos expresar $\frac{z}{w}$ en notación binómica, entonces multiplicamos y dividimos por el conjugado del denominador. De esta manera, en el denominador resulta $w\overline{w}$ que es real, y por lo tanto podemos separar parte real e imaginaria del cociente.
Veamos un ejemplo:

Ejemplo 1.13

$$z = \frac{3+2i}{1-i} = \frac{(3+2i)(1+i)}{(1-i)(1+i)} = \frac{3+3i-2+2i}{|1-i|^2} = \frac{1+5i}{2} = \frac{1}{2} + \frac{5}{2}i.$$

En realidad, muy probablemente esta no sea la primera vez que nos enfrentamos a la noción de complejos conjugados. Cuando buscamos raíces de polinomios de segundo grado, aparece la expresión $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$. Cuando los coeficientes son tales que $b^2 - 4ac$ es negativo, lo que obtenemos son dos raíces complejas conjugadas: $\frac{-b+i\sqrt{4ac-b^2}}{2a}$ y $\frac{-b-i\sqrt{4ac-b^2}}{2a}$. Recíprocamente, si tomamos dos complejos conjugados, digamos $z_0 = a + bi$ y $\overline{z_0} = a - bi$, y consideramos el polinomio que los tiene como raíces, entonces resulta:

$$P(z) = (z - z_0)(z - \overline{z_0}) = z^2 - z(z_0 + \overline{z_0}) + z_0\overline{z_0} = z^2 - 2az + |z_0|^2 = z^2 - 2az + a^2 + b^2.$$

Es decir, un polinomio con coeficientes reales.

El siguiente ejercicio da una generalización de este hecho.

Ejercicio 1.14
Sea $P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$ un polinomio con coeficientes a_i reales.

1. *Demostrar que $P(\overline{z}) = \overline{P(z)}$.*
2. *Concluir que si el polinomio tiene una raíz compleja, entonces su conjugada también es raíz.*

Volvamos con algunos aspectos geométricos. Primero, observemos que si z tiene argumento φ, entonces su conjugado $\bar{z}$ tiene argumento $-\varphi$ (y el mismo módulo, ver Figura 1.5).

Tomemos ahora un número θ, que denotará un ángulo, y consideremos el complejo $z = \cos(\theta) + i\sin(\theta)$. El módulo de este complejo es $|z| = \sqrt{\cos^2(\theta) + \sin^2(\theta)} = \sqrt{1} = 1$. Es decir, es un complejo que está en la circunferencia de centro en el origen y de radio uno. Recíprocamente, para cualquier complejo $z = a + bi$ de módulo uno (es decir, que $a^2 + b^2 = 1$), su parte real es $a = \cos(\theta)$ y su parte imaginaria $b = \sin(\theta)$, siendo θ un argumento de z.

Es decir, podemos recorrer la circunferencia unidad con números complejos de la forma $z = \cos(\theta) + i\sin(\theta)$, haciendo variar θ entre 0 y 2π.

Si ahora tomamos un complejo $z \neq 0$ cualquiera, y lo dividimos entre su módulo, el complejo resultante $\dfrac{z}{|z|}$ tiene módulo uno (y no es otra cosa que un re-escalado del complejo z, es como proyectarlo a la circunferencia unidad). Por lo tanto, como está en la circunferencia unidad, es $\dfrac{z}{|z|} = \cos(\theta) + i\sin(\theta)$, donde θ es un argumento de $\frac{z}{|z|}$ (o de z, pues tienen el mismo argumento).

Recuperamos entonces la expresión

$$z = |z|(\cos(\theta) + i\sin(\theta)) \tag{1.3}$$

que habíamos obtenido en (1.2). A esta notación de un complejo utilizando su módulo y argumento, se la denomina notación polar. A continuación vamos a definir la exponencial compleja, que entre otras cosas nos dará una notación polar más cómoda y compacta.

1.2. Exponencial compleja

La idea es buscar una función $f : \mathbb{C} \to \mathbb{C}$ que extienda a la función exponencial real que conocemos, y que además cumpla con las propiedades clásicas de la exponencial. Daremos una definición que puede parecer arbitraria, pero veremos que cumple con las propiedades deseadas.

Definición 1.15. Dado un complejo $z = a + bi$, definimos la función exponencial compleja como $e^z = e^a(\cos(b) + i\sin(b))$, donde e^a es la exponencial real conocida.

Lo primero que podemos observar es que si $z = a$ (es decir, con parte imaginaria nula), entonces $e^z = e^a(\cos(0) + i\sin(0)) = e^a$, por lo tanto en lo reales la exponencial recién definida coincide con la exponencial real.

Proposición 1.16. Si $z = a+bi$ y $w = c+di$ son dos complejos, entonces se tiene $e^{z+w} = e^z e^w$.

Demostración. Calculemos por separado e^{z+w} y $e^z e^w$, y veamos que son iguales.

Como $z+w = (a+c)+(b+d)i$, entonces

$$e^{z+w} = e^{a+c}(\cos(b+d)+i\sin(b+d)) = e^a e^c(\cos(b+d)+i\sin(b+d)).$$

Donde utilizamos la definición de la exponencial compleja, y la propiedad de la exponencial real $e^{a+c} = e^a e^c$.

Calculemos ahora $e^z e^w$. Esto es:

$$\begin{aligned} e^z e^w &= \big(e^a(\cos(b)+i\sin(b))\big)\cdot\big(e^c(\cos(d)+i\sin(d))\big) \\ &= e^a e^c\left[\underbrace{\cos(b)\cos(d)-\sin(b)\sin(d)}_{\cos(b+d)} + i\underbrace{(\cos(b)\sin(d)+\cos(d)\sin(b))}_{\sin(b+d)}\right] \\ &= e^a e^c(\cos(b+d)+i\sin(b+d)), \end{aligned}$$

donde fueron utilizadas las fórmulas para el seno y coseno de la suma. □

Comentario

Un comentario importante: no se pretende que estas fórmulas (del coseno y seno de la suma) sean recordadas de memoria. Es más, esta propiedad de los complejos es útil para *deducir* estas fórmulas. Es decir, ahora sabemos que $e^{i(\alpha+\beta)} = e^{i\alpha}e^{i\beta}$, y a partir de la definición, haciendo cuentas, llegamos a las fórmulas deseadas (es, en realidad, la demostración que recién hicimos, pero ahora la usamos "al revés": sabemos que el resultado es cierto, y lo aprovechamos para recordar las fórmulas que utilizamos para demostrarlo). ¡Hágalo!

Ejemplo 1.17
$e^{i\pi} = e^0(\cos(\pi)+i\sin(\pi)) = -1.$

Observemos que si z es un complejo puramente imaginario, $z = ib$, entonces $e^z = e^{ib} = \cos(b)+i\sin(b)$. Esto es, un complejo de módulo uno (y ángulo b).

En general, si $z = a+bi$, entonces el módulo de e^z es

$$|e^a(\cos(b)+i\sin(b))| = \sqrt{(e^a\cos(b))^2+(e^a\sin(b))^2} = e^a\sqrt{(\cos(b))^2+(\sin(b))^2} = e^a.$$

El módulo de e^z depende solamente de la parte real de z. Por otro lado, el argumento depende solamente de la parte imaginaria.

Otra forma de ver esto era observando que $e^{a+bi} = e^a e^{ib}$, de donde

$$|e^{a+ib}| = |e^a| \cdot |e^{ib}| = e^a.$$

Volvamos a la exponencial de un imaginario puro: $e^{ib} = \cos(b) + i\sin(b)$, y recordemos la notación polar vista en (1.3). Entonces, si z es un complejo de módulo ρ y argumento θ, tenemos la notación más compacta:

$$z = \rho e^{i\theta}.$$

Es decir, con $e^{i\theta}$ tenemos un complejo de módulo uno y ángulo θ. Si luego lo multiplicamos por un escalar $\rho > 0$, obtenemos un complejo de módulo ρ y ángulo θ, pues lo único que estamos haciendo al multiplicar por un real positivo, es justamente cambiar la escala.

Ejercicio 1.18
¿Cuál es la imagen por la función exponencial del conjunto de números complejos con parte real nula? ¿Y del conjunto de complejos con parte imaginaria nula?

Ejemplos 1.19

1. $1 = e^{i0}$.
2. $i = e^{i\frac{\pi}{2}}$.
3. $-i = e^{i\frac{3\pi}{2}}$.
4. $1 + i = \sqrt{2}e^{i\frac{\pi}{4}}$.

Ya vimos que la suma de complejos es muy natural, sobretodo si miramos su representación geométrica en el plano. Ahora, la interpretación para el producto de dos complejos no es evidente a partir de su definición. Sin embargo, si escribimos los números en su notación polar, $z_1 = \rho_1 e^{i\theta_1}$ y $z_2 = \rho_2 e^{i\theta_2}$, entonces el producto es:

$$z_1 z_2 = \rho_1 e^{i\theta_1} \cdot \rho_2 e^{i\theta_2} = \rho_1 \rho_2 e^{i(\theta_1+\theta_2)}.$$

O sea, cuando hacemos el producto de dos complejos, los módulos se multiplican, y los ángulos se suman.

Además de dar una interpretación geométrica clara, esta notación es muy cómoda para trabajar con productos y potencias de complejos. Veremos ejemplos de esto a continuación, y en la sección 1.3. En general:

Importante

La notación binómica es más cómoda para trabajar con sumas, y la notación polar es más cómoda para trabajar con productos y potencias.

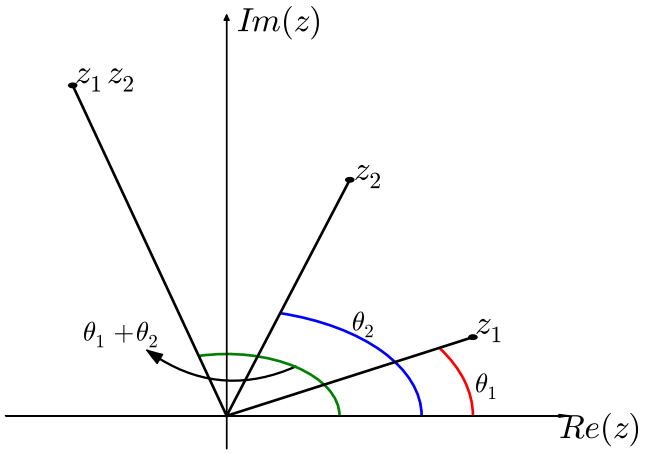


Figura 1.6: Producto de dos complejos. Se multiplican los módulos y se suman los ángulos.

Ejemplo 1.20
Busquemos todos los números $z \in \mathbb{C}$ que cumplan $z^2 = \bar{z}$.

Si escribimos z en su notación polar $z = \rho e^{i\theta}$, entonces buscamos que

$$\rho^2 e^{i2\theta} = \rho e^{-i\theta}.$$

Para que estos dos complejos sean iguales, deben serlo sus módulos y sus ángulos (a menos de múltiplos de 2π). Entonces tenemos:

$$\begin{aligned} &\rho^2 = \rho, \\ &2\theta = -\theta + 2k\pi, \quad \text{con } k \in \mathbb{Z}. \end{aligned}$$

De la primera ecuación, tenemos dos soluciones: $\rho = 0$ y $\rho = 1$.

De la segunda, tenemos que $\theta = \dfrac{2k\pi}{3}$, con $k \in \mathbb{Z}$. A priori podría parecer que tenemos infinitas soluciones (una para cada entero k). Sin embargo, rápidamente vemos que los ángulos se comienzan a repetir. Veamos qué sucede con algunos valores de k:

$$\begin{aligned} &\text{Si } k = 0 \quad \Longrightarrow \quad \theta_0 = 0, \\ &\text{Si } k = 1 \quad \Longrightarrow \quad \theta_1 = \frac{2\pi}{3}, \\ &\text{Si } k = 2 \quad \Longrightarrow \quad \theta_2 = \frac{4\pi}{3}, \\ &\text{Si } k = 3 \quad \Longrightarrow \quad \theta_3 = \frac{6\pi}{3} = 2\pi. \end{aligned}$$

Es decir, para $k = 3$ se repite el mismo ángulo que para $k = 0$, y si seguimos, para $k = 4$ se repetirá el mismo ángulo que para $k = 1$, y así. Tenemos por lo tanto cuatro soluciones:

$$z_0 = 0, \qquad z_1 = 1,$$
$$z_2 = e^{\frac{2\pi}{3}}, \qquad z_3 = e^{\frac{4\pi}{3}},$$

que se encuentran distribuidas según la siguiente figura:

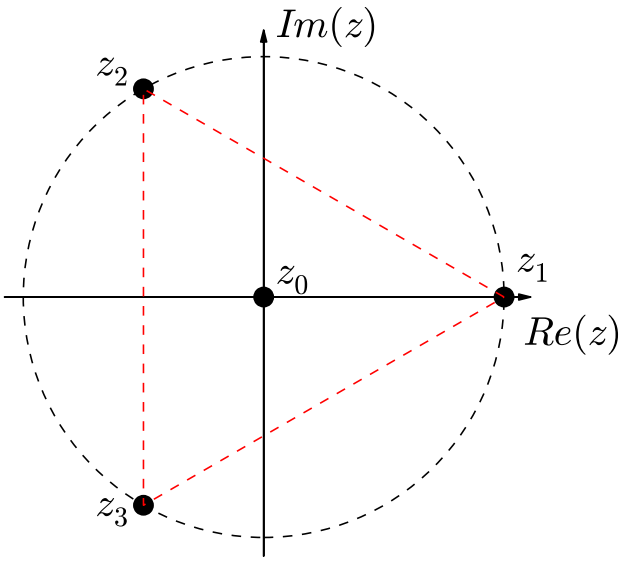


Figura 1.7: Las 4 soluciones de $z^2 = \bar{z}$.

¿Qué figura geométrica dibujan las soluciones distintas de cero? ¿Dónde se ve eso en las expresiones de las soluciones?

Ejercicio 1.21
Resuelva el mismo problema del ejemplo anterior, pero ahora utilizando la notación binómica, y verifique que las soluciones coinciden. ¿Qué pasaría con esta manera de proceder si en vez de z^2 tuviéramos z^5? ¿Y con la notación polar?

Observación 1.22
Observar que la función exponencial no es inyectiva (como sí lo es la exponencial en $\mathbb{R}$). Por ejemplo, tenemos que $e^{1+i\pi} = e^{1+i3\pi}$, y en general cualquier diferencia de múltiplos enteros de 2π en la parte imaginaria, dará el mismo resultado por la exponencial. Por lo tanto definir el logaritmo complejo como la función inversa de la exponencial no es viable. Se puede elegir un argumento entre $-\pi$ y π, y eso determina un logaritmo principal, pero no entraremos en eso en este curso.

1.3. Raíces complejas

Si bien hemos trabajado mucho con la unidad imaginaria, que verifica $i^2 = -1$, hemos tenido cuidado en no escribir $i = \sqrt{-1}$. Porque, veamos qué sucede si escribimos $i = \sqrt{-1}$ y operamos con las reglas que estamos acostumbrados:

$$-1 = i^2 = i \cdot i = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{1} = 1.$$

Es claro que hay algo mal. A esta altura, hay una observación que vale la pena hacer: cuando escribimos $i^2 = -1$ o $\sqrt{-1}$, no nos referimos al -1 como número real, sino al -1 como número complejo, como par ordenado $(-1,0)$ (que es cierto, lo *identificamos* con el -1 real, pero lo construimos como par ordenado, con sus propias operaciones). Y aún no hemos definido la raíz como operación dentro de los complejos, por lo que no sabemos qué propiedades tiene. ¿Será cierto por ejemplo que $\sqrt{zw} = \sqrt{z}\sqrt{w}$? Esta propiedad mantuvo a la comunidad matemática confundida por varios años[4].

Comencemos estudiando lo que se denominan las raíces enésimas de la unidad. Es decir, dado un natural n, buscamos números complejos z tales que $z^n = 1$. Entonces escribimos a z en su notación polar $z = \rho e^{i\theta}$, y tenemos que

$$\rho^n e^{in\theta} = 1.$$

De donde $\rho^n = 1$, y por lo tanto $\rho = 1$ (recordemos que ρ es un real no negativo, y por lo tanto $\rho = 1$ es la única solución). Luego, tenemos que $n\theta = 2k\pi$, con $k \in \mathbb{Z}$. De manera similar que en el ejemplo 1.20, de aquí obtenemos n soluciones, que son $\theta_k = \dfrac{2k\pi}{n}$. Observemos que para $k = 0$ obtenemos la solución $z_0 = 1$, que coincide con la raíz real conocida, y si n es par, para $k = \frac{n}{2}$ obtenemos $\theta_k = \pi$, es decir, $z_k = 1 \cdot e^{i\pi} = -1$.

Para dos valores consecutivos de k, las soluciones correspondientes difieren en un ángulo de $\frac{2\pi}{n}$, y todas tienen el mismo módulo. Es decir, tenemos n raíces enésimas de la unidad, y se encuentran en los vértices de un polígono regular de n lados.

Cuando queremos calcular las raíces enésimas de un complejo w_0 cualquiera, procedemos igual. Buscamos entonces los $z \in \mathbb{C}$ tales que $z^n = w_0$. Escribimos a w_0 en su notación polar $w_0 = \rho_0 e^{i\theta_0}$, y también al z que buscamos, $z = \rho e^{i\theta}$. Entonces tenemos:

$$\rho^n e^{in\theta} = \rho_0 e^{i\theta_0}.$$

De donde obtenemos dos ecuaciones reales, una para el módulo, y otra para el argumento:

$$\begin{aligned} \rho^n &= \rho_0, \\ n\theta &= \theta_0 + 2k\pi, \quad \text{con } k \in \mathbb{Z}. \end{aligned}$$

[4]De hecho, Euler en su libro sobre *Algebra* (alrededor de 1770) escribía $\sqrt{-1}\sqrt{-4} = \sqrt{4} = 2$ como un resultado posible.

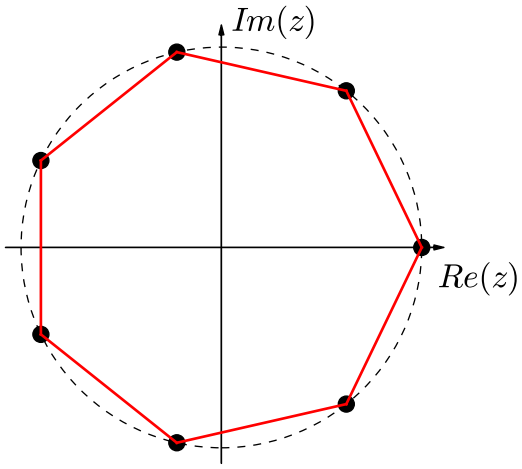


Figura 1.8: Complejos z que verifican $z^7 = 1$.

Recordando que ρ debe ser un real no negativo, y que las soluciones para θ son cíclicas en k, obtenemos:

$$\rho = \sqrt[n]{\rho_0},$$
$$\theta = \frac{\theta_0}{n} + \frac{2k\pi}{n}, \quad k = 0, \ldots, n-1.$$

Nuevamente tenemos n soluciones, todas de igual módulo, y formando un polígono regular. La diferencia con las raíces enésimas de la unidad son dos: el módulo no tiene por qué ser uno, naturalmente, y el "primer" ángulo (el correspondiente a $k = 0$) no es $\theta = 0$ sino el ángulo de w_0 dividido n.

Nota histórica

En el siglo XVI la comunicación y publicación de resultados, en particular en matemática, no se daba de igual manera que ahora. Eran comunes los desafíos, en los que matemáticos se proponían problemas y se retaban a ver quién era capaz de resolver más instancias, usualmente con un premio monetario, o con plazas en las universidades de por medio.

La historia completa sobre lo que hoy se conoce como la fórmula de *Cardano-Tartaglia* (o de *del Ferro-Tartaglia-Cardano*) para hallar soluciones a ecuaciones cúbicas, se conoce parcialmente.

La primera solución es debida a Scipione del Ferro (Bologna, 1465 - 1526), quien guardaba sus descubrimientos en un cuaderno que no compartía con nadie, para tener ventajas en caso de desafíos. Le comunicó esta solución a un estudiante suyo llamado Antonio Fiore, en su lecho de muerte según algunas historias.

Algunos años más tarde, Niccolo "Tartaglia" Fontana (apodado así debido a su tartamudez, causada por una cuchillada de un soldado francés durante la masacre de 1512 en Brescia) también se encontraba trabajando en las soluciones a ecuaciones cúbicas, y fue desafiado por Fiore. El reto consistió en 30 problemas que cada uno le proponía al contrincante, y fue Tartaglia quien ganó ese duelo, debido a que su fórmula funcionaba en más casos.

Entonces apareció Gerolamo Cardano, un matemático que ya contaba con cierta reputación. Tartaglia visita a Cardano en Milan en 1539, y le confía su fórmula, después de ser persuadido por Cardano. De hecho Tartaglia considera que fue presionado a entregarla a cambio de favores políticos (para conseguir una posición en una Universidad). La fórmula se la da en forma de poema, por si llegaba a caer en manos extrañas, y bajo el juramento de Cardano de no publicarla.

En 1543, Cardano y su estudiante Ludovico Ferrari viajan a Bologna a encontrarse con el yerno de Scipione del Ferro, Hannival Nave, quien había heredado su cuaderno de anotaciones, donde aparecía la solución a la ecuación cúbica.

Como la solución de del Ferro precedía a la de Tartaglia, Cardano decidió que no estaría violando su juramento hecho a Tartaglia de conservar en secreto su fórmula, y publicó la solución en su libro *Ars Magna* en 1545, pero reconociendo la autoría de Scipione del Ferro y de Tartaglia.

A Tartaglia de todas maneras no le gustó esto, comenzando así un extenso intercambio de insultos y desafíos con Ferrari, el estudiante de Cardano.

Capítulo 2

Sucesiones y Series

2.1. Sucesiones

En esta primera parte del capítulo nos vamos a ocupar de las sucesiones, que son un objeto muy natural e intuitivo, pero que de todas maneras presentan particularidades interesantes y delicadas. Las sucesiones tienen uso en sí mismo para modelar aplicaciones de todo tipo, pero además sirven como herramienta dentro de la matemática, para caracterizar objetos (continuidad o conjuntos por ejemplo) y para demostrar resultados.

Intuitivamente, una sucesión es una lista ordenada e infinita de números reales, que se pueden repetir:

Definición 2.1. Una sucesión es una función de los naturales en los reales, $a : \mathbb{N} \to \mathbb{R}$. Al elemento enésimo de la sucesión (es decir, $a(n)$) se lo denota a_n, y a la sucesión entera $(a_n)_{n\in\mathbb{N}}$.

Muchas veces hacemos un abuso de notación y nos referimos a la sucesión como a_n simplemente. Quedará claro por el contexto cuándo nos referimos a la sucesión como objeto, y cuándo a un término específico de la misma.

De la definición como función $a : \mathbb{N} \to \mathbb{R}$, lo realmente importante es que el dominio sean los naturales. Si cambiamos a los reales por otro conjunto, tendremos otros tipos de sucesiones, por ejemplo, de números complejos, o de puntos en $\mathbb{R}^n$, como estudiaremos en los capítulos siguientes. Pero lo que captura la idea de lista ordenada es justamente la asignación de un elemento para cada lugar n de la lista infinita.

Si bien la definición de sucesión es como una función, la naturaleza de las mismas hace que sean muy distintas a las funciones que fueron objeto de estudio del primer curso. Es decir, la mayoría de los conceptos importantes que fueron estudiados para funciones definidas en los reales (límite de una función cuando x tiende a un punto a, continuidad, derivabilidad), se

basan fuertemente en que uno se puede acercar tanto como quiera a un punto a. Esto carece de sentido cuando el dominio son los números naturales, donde el único límite que podemos tomar, como ya veremos, es cuando el índice n tiende a infinito.

Ejemplos 2.2

1. $a_n = \frac{1}{n}$, *sus elementos son:* $\left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right)$.
2. $a_n = 1$, $\forall n$.
3. $a_n = (-1)^n$.
4. $a_n = n$.

Vale la pena también advertir que no se debe confundir a la sucesión en sí, con el recorrido de la misma, que es simplemente la imagen como función. Por ejemplo, la segunda sucesión de los ejemplos es $(1,1,1,\ldots)$, y la imagen de la sucesión es el conjunto $\{1\}$, formado por un solo elemento. En el siguiente ejemplo, el recorrido de $a_n = (-1)^n$ es el conjunto $\{-1,1\}$, y la sucesión es $(-1,1,-1,1,\ldots)$.

En general, el comportamiento que más nos va a interesar es lo que pase con la sucesión "en el infinito"[1]. Vamos a decir que una sucesión tiene límite L (o también que *converge a* L o *tiende a* L) cuando, para cualquier entorno de L (o sea, tan cerca como queramos estar de L), existe un momento a partir del cual la sucesión se encuentra completamente en ese entorno. Formalmente:

Definición 2.3. Decimos que la sucesión a_n tiene límite $L \in \mathbb{R}$, y lo denotamos $\lim_{n\to\infty} a_n = L$ sii $\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}$ tal que $\forall n \geqslant n_0$ se tiene que $a_n \in E(L,\varepsilon)$[2].

Una manera de pensarlo:

Es un "juego" de $\varepsilon - n_0$ similar al de continuidad con $\varepsilon - \delta$. Es decir, un jugador va eligiendo valores de ε para que el segundo jugador conteste con un momento n_0 a partir del cual $a_n \in E(L,\varepsilon)$. Cuando vayamos a demostrar que una sucesión a_n tiene límite, usando la definición, nos pondremos en el papel del segundo jugador, y tendremos que encontrar el n_0 para un ε dado. Si, en otro caso, sabemos que determinada sucesión tiene límite, y pretendemos usar la definición, entonces estaremos en el papel del primer jugador. Más adelante comentaremos sobre esta dualidad con casos específicos.

[1] Ejercicio avanzado: demostrar que la definición de límite no depende del orden de los elementos de la sucesión. Es decir, si $\varphi : \mathbb{N} \to \mathbb{N}$ es una biyección, entonces $\lim_n a_n = \lim_n a_{\varphi(n)}$, si existe.

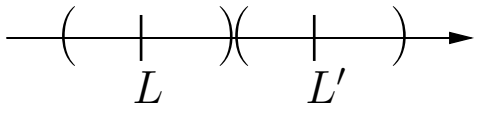
[2] Recordamos que el entorno de centro L y radio ε es el conjunto de puntos $x \in \mathbb{R}$ tales que $|x - L| < \varepsilon$.

En el ejemplo $a_n = \dfrac{1}{n}$, a medida que n crece, su inverso es cada vez más chico. Demostremos que efectivamente tenemos $\lim \dfrac{1}{n} = 0$. Para esto, según la definición, debemos probar que para todo ε existe un natural n_0 a partir del cual la sucesión está a menos de ε de cero. La manera de demostrar que para todo ε se cumple una condición, es tomar un ε genérico, y, en este caso, encontrar un n_0 que sirva. Tomemos entonces un $\varepsilon > 0$ genérico. Lo que queremos es que $|a_n - 0| < \varepsilon$, o sea, que $\dfrac{1}{n} < \varepsilon$. Esto úlimo se cumplirá siempre que n sea mayor que $\dfrac{1}{\varepsilon}$, por lo que alcanza con tomar como n_0 cualquier natural mayor que $\dfrac{1}{\varepsilon}$.

Para el ejemplo de la sucesión constante $a_n = 1$, resulta claro que el límite debe ser uno. Dado entonces un ε cualquiera, ¿cuál es el n_0 que sirve? ¿Qué pasa con la sucesión $a_n = n$?

Proposición 2.4. Si existe el límite de a_n, entonces es único.

Demostración. Supongamos que a_n tiene dos límites, $L \neq L'$, y tomemos $\varepsilon = \dfrac{|L - L'|}{2}$, de manera que $E(L,\varepsilon) \cap E(L',\varepsilon) = \emptyset$.



A partir de la definición de límite para L, tenemos que existe un $n_1 \in \mathbb{N}$ tal que $a_n \in E(L,\varepsilon)$ para todo $n \geqslant n_1$. Como supusimos que L' también es límite, de la misma forma tenemos que existe un n_2 tal que $a_n \in E(L',\varepsilon)$ para todo $n \geqslant n_2$. Pero entonces a partir de $n_3 = \text{máx}\{n_1, n_2\}$, la sucesión debe estar en los dos entornos al mismo tiempo, cosa que es imposible ya que los entornos son disjuntos. Por lo tanto llegamos a un absurdo, de donde $L = L'$. □

Definición 2.5. Decimos que la sucesión a_n está acotada sii $\exists K \in \mathbb{R}$ tal que $|a_n| \leqslant K, \forall n \in \mathbb{N}$.

También decimos que la sucesión está acotada superiormente cuando existe un $K \in \mathbb{R}$ tal que $a_n < K$ para todo $n \in \mathbb{N}$, y de manera similar se define la acotación inferior. Veamos que las sucesiones convergentes están acotadas.

Proposición 2.6. Si a_n tiene limite, entonces está acotada.

Demostración. Tomemos un ε cualquiera, por ejemplo $\varepsilon = 1$. Como a_n es convergente, a partir de un $n_0 \in \mathbb{N}$, la sucesión está completamente comprendida dentro de $E(L,1)$. Tenemos entonces que $|a_n| < |L| + 1$ para $n \geqslant n_0$. Los primeros términos (antes del n_0), no sabemos dónde están, pero son una cantidad finita, por lo que hay alguno que será el más grande de todos en valor absoluto. Si tomamos $K = \text{máx}\{|a_1|, |a_2|, \ldots, |a_{n_0-1}|, |L| + 1\}$, entonces tenemos que $|a_n| < K$ para todo $n \in \mathbb{N}$. □

El recíproco de este resultado no es cierto, y la sucesión $a_n = (-1)^n$ sirve como contra-ejemplo. Está acotada (se puede tomar cualquier $K \geqslant 1$), pero no tiene límite (¿qué ε tomaría para demostrar que no converge a 1 por ejemplo?).

Por otro lado, la sucesión del ejemplo $a_n = n$, claramente no está acotada, ni tiene límite finito. Por el contrario, a_n es cada vez más grande, y de alguna manera "se acerca a infinito".

Definición 2.7. Decimos que el límite de a_n es infinito, $\lim_{n\to\infty} a_n = \infty$, sii $\forall K > 0, \exists n_o \in \mathbb{N}$ tal que $a_n > K$ para todo $n \geqslant n_0$.

Así, tenemos que para $a_n = n$, $\lim a_n = \infty$ (dado un K, ¿cuál es el n_0 que sirve para la definición?). De manera similar, que queda como ejercicio, se define $\lim\limits_{n\to\infty} a_n = -\infty$. Cuando una sucesión no tiene límite finito ni infinito, decimos que oscila.

Tenemos entonces que las sucesiones acotadas pueden ser convergentes o no. ¿Qué pasa con las sucesiones no acotadas? ¿Necesariamente tienen límite más o menos infinito? Estudiar acotación y convergencia de $a_n = (-1)^n n$.

Definición 2.8. Decimos que una sucesión a_n es mónotona creciente sii $a_{n+1} \geqslant a_n, \forall n \in \mathbb{N}$, y que es monótona decreciente sii $a_{n+1} \leqslant a_n, \forall n \in \mathbb{N}$. Cuando la desigualdad es estricta, decimos que la sucesión es estrictamente monótona.

Teorema 2.9. Si a_n es una sucesión monótona creciente y acotada superiormente, entonces tiene límite.

Demostración. Sea $L = sup\{a_n : n \in \mathbb{N}\}$, es decir, el supremo del conjunto imagen de la sucesión. Como este conjunto es acotado, por el axioma de completitud, este supremo existe. Ahora, dado cualquier $\varepsilon > 0$, tiene que haber algún elemento de la sucesión en $(L-\varepsilon, L]$, pues de otra manera $L - \varepsilon$ sería cota superior. Llamemos a_{n_0} a ese elemento. Luego, como a_n es creciente, todos los elementos posteriores son mayores que a_{n_0}. Por lo tanto, a partir de ese n_0, $a_n \in (L-\varepsilon, L]$, como queríamos probar. □

De forma análoga tenemos que toda sucesión monótona decreciente y acotada inferiormente, tiene límite.

Los siguientes ejercicios, además de ser resultados necesarios, ponen en práctica las técnicas de demostración que venimos desarrollando en este capítulo.

Ejercicio 2.10

1. *Demostrar que si* $a_n \leqslant b_n, \forall n \in \mathbb{N}$ *y ambas tienen límite, entonces* $\lim a_n \leqslant \lim b_n$.
2. *Si* a_n *es acotada, y* $\lim b_n = 0$*, demostrar que* $\lim a_n b_n = 0$.

Veamos ahora algunas propiedades algebraicas de los límites de sucesiones, muy similares a los vistos para límites de funciones.

Teorema 2.11. Sean a_n y b_n dos sucesiones con $\lim_{n\to\infty} a_n = A$ y $\lim_{n\to\infty} b_n = B$. Entonces:

1. $\lim_{n\to\infty} a_n + b_n = A + B$.
2. $\lim_{n\to\infty} a_n - b_n = A - B$.
3. $\lim_{n\to\infty} a_n b_n = AB$.
4. Si $B \neq 0$, entonces $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{A}{B}$.

Demostración.

1. Buscamos acotar $|a_n + b_n - A - B|$. Por la desigualdad triangular, tenemos que $|a_n + b_n - A - B| \leqslant |a_n - A| + |b_n - B|$, y cada uno de esos términos lo podemos hacer tan chico como queramos. Específicamente, sea $\varepsilon > 0$. Entonces, como $\lim a_n = A$, existe un $n_1 \in \mathbb{N}$ a partir del cual $|a_n - A| < \varepsilon/2$.Análogamente, tenemos que existe un $n_2 \in \mathbb{N}$ a partir del cual $|b_n - B| < \varepsilon/2$. Tenemos por lo tanto que a partir de $n_0 = \max\{n_1, n_2\}$, se cumple $|a_n + b_n - A - B| < \varepsilon$ como queríamos probar.
2. Análogo al caso anterior.
3. Se basa en lo siguiente (sumando y restando un término adecuado): $|AB - a_n b_n| = |AB - Ab_n + Ab_n - a_n b_n| \leqslant |A||B - b_n| + |b_n||A - a_n|$.
4. Basta demostrar que $\lim \frac{1}{b_n} = \frac{1}{B}$, y utilizar la propiedad del producto. Para esto, observar que $\left|\frac{1}{B} - \frac{1}{b_n}\right| = \frac{|B - b_n|}{|B||b_n|}$, y acotar inferiormente $|b_n|$ por $|B|/2$ (por ejemplo).

□

Ejercicio 2.12
Completar las demostraciones del teorema anterior.

Volviendo a los jugadores:

Detengámonos a ver cómo usamos la definición de límite, en términos de los dos jugadores que mencionamos más arriba. En la unicidad del límite (Prop. 2.4), jugamos (dos veces) en el papel del segundo jugador: elegimos un ε conveniente para nosotros, y la definición nos devuelve un momento n_1 o n_2. Algo similar hicimos en la Proposición 2.6. Cuando demostramos que las sucesiones monótonas y acotadas tienen límite (Teorema 2.9), nos pusimos del otro lado: dado un ε genérico, nosotros encontramos el n_0.

Quizás la primera de las propiedades que vimos recién, la del límite de la suma, es la que mejor muestra esta dualidad. Tenemos dos sucesiones (a_n y b_n) que sabemos que tienen

límite, y una tercera (la suma $a_n + b_n$) que queremos demostrar que lo tiene. Para esta sucesión suma entonces, debemos tomar el ε dado y encontrar un n_0. Tomamos este ε, lo dividimos convenientemente entre dos, y, desde el papel del jugador dos, para ese valor $\frac{\varepsilon}{2}$ que ahora nosotros damos, recibimos un natural para cada sucesión, a_n y b_n. Con estos, ahora, devolvemos un natural n_0 que cumple lo pedido.

Observar que las propiedades recién enunciadas son ciertas para límites finitos. Algunas de ellas se pueden extender también cuando los límites son infinitos, o cuando uno es finito y el otro infinito. Sin embargo, hay que tener más cuidado en estos casos. Por ejemplo, si $a_n \to \infty$ y $b_n \to -\infty$, no podemos afirmar nada de la suma $a_n + b_n$. Se pueden encontrar ejemplos (búsquelos) donde la suma tiende a $+\infty$, a $-\infty$, a cualquier real, o que no tenga límite. Estos casos se denominan *indeterminados*, y también sucede con productos y potencias, como por ejemplo $0 \times \infty$, $0/0$, ∞/∞, ∞^0, 1^∞, y 0^0.

En definitiva, tenemos el mismo comportamiento que se explica en las proposiciones 56 y 59 del capítulo *Límites y continuidad* de las notas del curso [8].

Ejercicio 2.13
Decimos que un par de sucesiones a_n y b_n forman un Par de Sucesiones Monótonas Convergentes (PSMC) sii

(i) *a_n es creciente y b_n es decreciente.*

(ii) *$a_n \leqslant b_n$ para todo $n \in \mathbb{N}$.*

(iii) *Dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que $b_{n_0} - a_{n_0} < \varepsilon$.*

Entonces:

1. *Demostrar que $a_n \leqslant b_m$ para todo $n, m \in \mathbb{N}$.*
2. *Deducir que ambas sucesiones tienen límite, que llamaremos $\lim a_n = L$ y $\lim b_n = L'$.*
3. *Deducir que $L \leqslant L'$.*
4. *Demostrar que $L = L'$ (y observar que recién ahora es necesaria la propiedad (iii)).*

Ejercicio 2.14
Sean $I_n = [a_n, b_n]$ una sucesión de intervalos tales que $I_0 \supset I_1 \supset I_2 \supset \cdots \supset I_n \supset \cdots$, y además la longitud del intervalo tiende a cero con n. Demostrar que las sucesiones de extremos de los intervalos a_n y b_n forman un PSMC, y que si L es su límite, entonces $\bigcap_{n=1}^{\infty} I_n = L$.

Ejemplo 2.15
En 1683, el matemático suizo Jacob Bernoulli estaba estudiando ciertas cuestiones de interés compuesto[3]*. Supongamos que depositamos un peso en una cuenta con un 100 % de interés*

[3] En el interés compuesto, las ganancias por el interés se van agregando al monto sobre el cual se va calculando el interés.

anual. Si el interés se acredita una única vez, entonces al final del año la cuenta tendrá dos pesos. Si se acredita dos veces, entonces al pasar los primeros seis meses, se acreditará el 50% *(se multiplica por* 1,5*), y luego sobre esos* \$1,50 *se calcula nuevamente el* 50%*, lo que da un total de* $1{,}00 \times (1{,}5)^2 = 2{,}25$. *Si dividimos el* 100% *de interés en cuatro, acreditando el* 25% *cada cuatro meses, tendremos al final del año* $1{,}00 \times (1{,}25)^4 = 2{,}4414\ldots$. *Es intuitivo, que cuanto más frecuente se hacen las acreditaciones, mayor será el saldo a final del año. En general, si dividimos en n intervalos, el saldo final será* $1{,}00 \times \left(1+\frac{1}{n}\right)^n$. *¿Qué pasa cuando n crece?*

Si desarrollamos la expresión según el denominado binomio de Newton, obtenemos n sumandos, y cada uno crece con n, por lo tanto la sucesión $a_n = \left(1+\frac{1}{n}\right)^n$ *es creciente. Del mismo desarrollo se puede observar también que es acotada, y por lo tanto tiene límite. El límite es exactamente el número e (de hecho es una de las definiciones más usadas para e).*

Calcular límites es una de las tareas más importantes, pero recurrir a la definición cada vez, puede resultar tedioso, sobre todo para sucesiones complicadas. Por lo tanto muchas veces lo que hacemos es estudiar sucesiones que se comportan igual en el infinito, pero que son mucho más fáciles de estudiar. El concepto de sucesiones equivalentes formaliza esto:

Definición 2.16. Decimos que dos sucesiones a_n y b_n, ambas con límite 0 o ∞, son equivalentes sii $\lim_{n\to\infty} \dfrac{a_n}{b_n} = 1$.

Como ya es sabido por el lector, hay que tener cuidado con la sustitución por sucesiones equivalentes, cuando resulta en la resta de dos equivalentes que se anulan. Ejemplos donde la cancelación indiscriminada de términos equivalentes resulta en errores, ya se ha visto con desarrollos de Taylor de funciones. Con sucesiones pasa lo mismo. Cuando dos sucesiones son equivalentes, se denota $a_n \sim b_n$.

Ejemplo 2.17
Las sucesiones $a_n = 4n^3+2n+1$ *y* $b_n = 4n^3$ *son equivalentes. En efecto,*

$$\lim_{n\to\infty} \frac{4n^3+2n+1}{4n^3} = \lim_{n\to\infty} \frac{4n^3}{4n^3} + \frac{1}{2n^2} + \frac{1}{4n^3} = 1.$$

Cuando tenemos un polinomio en n, es equivalente a su término de mayor grado. En general, tenemos las mismas equivalencias que teníamos para funciones, y que fueron refinadas con el desarrollo de Taylor. Hay que tener cuidado con lo siguiente: cuando calculamos límites de sucesiones, siempre lo hacemos con n tendiendo a infinito. Sin embargo, podemos tener por ejemplo $\operatorname{sin}\left(\dfrac{1}{n}\right)$. En este caso, cuando n tiende a infinito, $\dfrac{1}{n}$ tiende a cero, y entonces debemos utilizar el equivalente del seno en cero. Es decir, $\operatorname{sin}\left(\dfrac{1}{n}\right) \sim \dfrac{1}{n}$ cuando n tiende a infinito.

Ejercicio 2.18
Corroborar los siguientes límites

1. $\lim_{n\to\infty} n\left[\log(n+1)-\log(n)\right]=1$
2. $\lim_{n\to\infty} n^2 \sin\left(\frac{1}{n}\right)=\infty$
3. $\lim_{n\to\infty} n\left[n-\sqrt{n^2-1}\right]=1/2$

Volvamos a la sucesión $a_n=(-1)^n$. Ya vimos que no tiene límite, pero está acotada, y de hecho pasa infinitas veces por el -1 y por el 1. Más aún, si miramos solamente los términos pares (o impares), obtenemos una sucesión constante, y por lo tanto convergente. Las subsucesiones permiten generalizar este ejemplo, y obtener resultados importantes en general.

Intuitivamente, una subsucesión consiste en tomar *algunos* (infinitos) términos de la sucesión. Por ejemplo, los términos pares. La definición formal de este concepto tan simple es complicada de apariencia, pero la idea es esa: elegir infinitos elementos de la sucesión original.

Definición 2.19. Sea a_n una sucesión real, y n_k una sucesión estrictamente creciente de números naturales. Entonces la sucesión a_{n_k} es una subsucesión de a_n.

La sucesión n_k es la que juega el papel de elegir los índices con los que nos vamos a quedar, y al ser una sucesión creciente de naturales, tenemos asegurado que n_k tiende a infinito (y por lo tanto vamos a elegir infinitos términos de a_n).

Ejemplos 2.20

1. *Para la sucesión $a_n=(-1)^n$, podemos tomar la subsucesión de los pares $a_{2n}=1$, y de los impares $a_{2n+1}=-1$. ¿Qué pasa con a_{3n}? ¿Cuántas subsucesiones convergentes tiene a_n?*
2. *Para la sucesión $a_n=n+(-1)^n n$, tenemos por ejemplo $a_{2n+1}=0$. ¿Qué pasa con a_{2n}?*

En los ejemplos de arriba, encontramos subsucesiones convergentes, pero este no es siempre el caso. Intente encontrar una sucesión que no tenga ninguna subsucesión convergente.

Cuando la sucesión original es convergente, todas sus subsucesiones heredan el mismo límite:

Teorema 2.21. Si $\lim a_n=L$, entonces toda subsucesión de a_n converge a L.

Demostración. Sea a_{n_k} una subsucesión, y sea $\varepsilon>0$. Como $a_n\to L$, existe un $n_0\in N$ a partir del cual $a_n\in E(L,\varepsilon)$. Pero como n_k tiende a infinito, existe k_0 tal que $n_k>n_0$ para todo $k>k_0$, de donde $a_{n_k}\in E(L,\varepsilon)$ a partir de n_{k_0}. □

Daremos ahora una definición topológica, que involucra puntos y conjuntos pero no sucesiones, para luego sí, volver sobre las sucesiones. Ahondaremos mucho más sobre este tipo de definiciones y propiedades en el capítulo 4.

Definición 2.22. Decimos que un punto $a \in \mathbb{R}$ es un punto de acumulación de un conjunto $A \subset \mathbb{R}$ sii todo entorno reducido de a contiene algún punto de A. Es decir, si $\forall \varepsilon > 0$ se tiene $E^*(a,\varepsilon) \cap A \neq \emptyset$.

Observar que nada se dice sobre la pertenencia de a al conjunto. Es decir, puede haber puntos de acumulación que sean elementos de A, y puede haber puntos de acumulación de A que no estén en el conjunto.

Ejemplos 2.23

1. *Para el conjunto $A = [0,1)$, todos los puntos de $[0,1]$ son puntos de acumulación de A. Observar que el 1 es de acumulación (efectivamente cualquier entorno de 1 interseca al conjunto), pero no pertenece a A.*

2. *Para el conjunto $A = [0,1] \bigcup \{2\}$, los puntos de acumulación son los de $[0,1]$. Observar que 2, a pesar de estar en el conjunto, no es un punto de acumulación. En efecto, para serlo, debería haber puntos de A en cualquier entorno reducido de 2. Pero si tomamos por ejemplo $E^*(2,1/2)$, no hay ningún elemento de A en dicho entorno reducido.*

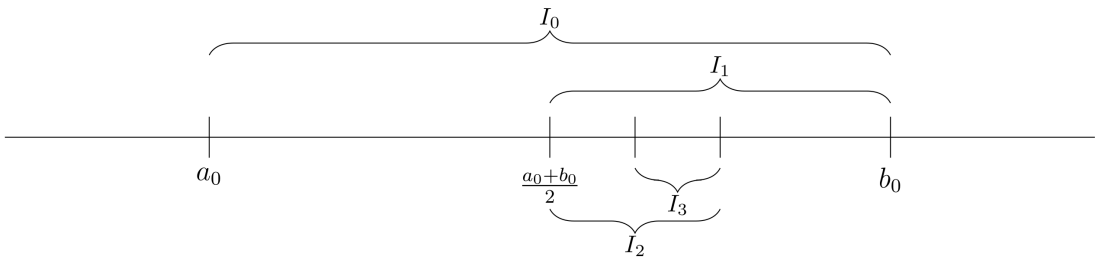
3. *Si $A = \left\{ \dfrac{1}{n} : n \in \mathbb{N} \right\}$, entonces el único punto de acumulación es el 0 (que además no está en el conjunto).*

4. *Si $A = \mathbb{Q}$, entonces, debido a la densidad de los racionales, todos los reales son puntos de acumulación de $\mathbb{Q}$.*

5. *Si $A = \mathbb{N}$, ¿cuáles son sus puntos de acumulación?*

Teorema 2.24 (Bolzano-Weierstrass)**.** Todo conjunto infinito y acotado tiene (al menos) un punto de acumulación.

Demostración. Llamemos A al conjunto. Como está acotado, lo podemos incluir en un intervalo, al que llamaremos I_0: $A \subset I_0 = [a_0, b_0]$. Ahora, si dividimos a este intervalo en dos partes iguales, $\left[a_0, \dfrac{a_0+b_0}{2}\right]$ y $\left[\dfrac{a_0+b_0}{2}, b_0\right]$, como el conjunto A es infinito, entonces (al menos) uno de las dos mitades debe contener infinitos puntos de A. Llamemos $I_1 = [a_1, b_1]$ a ese intervalo. Repitiendo el argumento, si dividimos I_1 en dos mitades, debe haber una de ellas con infinitos elementos de A, al que llamaremos I_2. Así, generamos una sucesión de intervalos encajados $I_{n+1} \subset I_n$, todos ellos con infinitos elementos de A, y la longitud de I_n es $\dfrac{b_0 - a_0}{2^n}$.



Por el ejercicio 2.13, la intersección de estos intervalos define un punto $L = \bigcap_{n\in\mathbb{N}} I_n$. Veamos que L es un punto de acumulación de A. Para esto, tomemos un $\varepsilon > 0$, y sea N natural tal que $\dfrac{b_0 - a_0}{2^N} < \varepsilon$. Entonces, como $L \in I_N$ y el tamaño de este intervalo es menor que el radio del entorno, tenemos que $I_n \subset E(L,\varepsilon)$. Luego, como en I_N hay infinitos puntos de A, hay al menos un punto de A en $E^*(L,\varepsilon)$ (es más, hay infinitos). □

Ejercicio 2.25
Para ver que ambas hipótesis son realmente necesarias, encontrar:

- *Un conjunto infinito (pero no acotado), que no contenga puntos de acumulación.*
- *Un conjunto acotado (pero no infinito), que no contenga puntos de acumulación.*

Volvamos a las sucesiones, aplicando el teorema recién demostrado para probar otro resultado sobre sucesiones, que también se conoce con el nombre de Bolzano-Weierstrass.

Teorema 2.26. Toda sucesión a_n acotada tiene una subsucesión convergente.

Demostración. Llamemos A al conjunto recorrido de la sucesión: $A = \{a_n : n \in \mathbb{N}\}$. Como la sucesión es acotada, el conjunto A es acotado. Puede ocurrir que A sea finito o infinito. Si A es finito, entonces la sucesión pasa infinitas veces por alguno de sus elementos. Tomando esos índices, construimos una subsucesión que converge a ese elemento. Si A es infinito, podemos aplicar el Teorema 2.24, y por lo tanto A tiene un punto de acumulación L. Construiremos una subsucesión de a_n convergente a L. Como L es de acumulación, en cualquier entorno habrá puntos de a_n. Tomemos entonces valores sucesivos de radios de los entornos: $\varepsilon_k = \dfrac{1}{k}$. En cada entorno $E^*(L,\varepsilon_k)$ hay un elemento de la sucesión, al que llamaremos a_{n_k}. Por construcción, $|a_{n_k} - L| < \dfrac{1}{k}$, por lo que $a_{n_k} \to L$. □

En el ejemplo que venimos repitiendo, $a_n = (-1)^n$, el recorrido es el conjunto $\{-1,1\}$, por lo que estamos en el primer caso (A finito) de la demostración anterior. La sucesión pasa infinitas veces por cada valor.

Queda como ejercicio encontrar una sucesión acotada que no sea convergente, y que caiga en el segundo caso de la demostración.

2.1.1. Sucesiones y completitud

Hasta ahora, trabajamos durante todo el capítulo con sucesiones reales, aunque comentamos que puede haber sucesiones en otros conjuntos. Cabe preguntarse cuáles de los resultados que vimos siguen siendo ciertos cuando, por ejemplo, consideramos las sucesiones definidas en $\mathbb{Q}$. Es decir, tomemos una sucesión $a : \mathbb{N} \to \mathbb{Q}$, y olvidemos que hay otros números reales que no son racionales. ¿Sigue siendo cierto que toda sucesión monótona y acotada es convergente? ¿Que toda sucesión acotada tiene una subsucesión convergente?

Tomemos por ejemplo la sucesión $a : \mathbb{N} \to \mathbb{Q}$ dada por $a_n = \left(1 + \frac{1}{n}\right)^n$. Claramente, para cualquier n natural, $1 + \frac{1}{n}$ es racional, y por lo tanto también lo es a_n. Sin embargo, esta sucesión no tiene límite en $\mathbb{Q}$.

Cuando construimos los números reales axiomáticamente, observamos que lo único que diferencia (desde el punto de vista de los axiomas) a $\mathbb{R}$ de $\mathbb{Q}$ es el Axioma de completitud. Entonces, si estos resultados no son ciertos para $\mathbb{Q}$, quiere decir que utilizamos el Axioma de completitud para su demostración. En la demostración del Teorema 2.9 lo hicimos explícitamente. Es un lindo ejercicio analizar qué resultados valen para sucesiones en $\mathbb{Q}$, y para los que no son ciertos, identificar dónde fue utilizado el Axioma de completitud.

Esta diferencia sustancial entre $\mathbb{Q}$ y $\mathbb{R}$, se denomina justamente completitud. Básicamente el conjunto de los racionales tiene "agujeros", es decir, sucesiones que se van apretando alrededor de algo que no está en el conjunto. Cabe preguntarse si por ejemplo el conjunto de números complejos es completo o no, pero lamentablemente el Axioma de completitud depende fuertemente de tener un cuerpo ordenado, cosa que no tenemos en $\mathbb{C}$. Lo que adelantaremos ahora, y profundizaremos en el capítulo 4, es una definición de completitud que vamos a poder aplicar en cualquier conjunto donde podamos definir sucesiones.

Definición 2.27. Decimos que una sucesión a_n es de Cauchy sii $\forall \varepsilon > 0$, existe un $n_0 \in \mathbb{N}$ tal que $|a_n - a_m| < \varepsilon$ para todo $n, m > n_0$.

Observar que la definición es muy similar a la definición de límite, pero en lugar de pedir que la sucesión esté cerca de L a partir de un momento, pedimos que se aprete entre sí. Es más, en la definición de sucesión de Cauchy solamente aparecen elementos de la sucesión, y no puntos externos como un candidato a límite L.

Los siguientes resultados relacionan la condición de Cauchy con la convergencia.

Teorema 2.28. Si a_n es una sucesión convergente, entonces es de Cauchy.

Demostración. Llamemos L al límite de a_n, y sea $\varepsilon > 0$. Como $a_n \to L$, entonces existe un $n_0 \in \mathbb{N}$, tal que $a_n \in E(L, \varepsilon/2)$ para todo $n \geqslant n_0$. Entonces si n y m son mayores a n_0, tenemos que $|a_n - L| < \varepsilon/2$ y $|a_m - L| < \varepsilon/2$, de donde $|a_n - a_m| < \varepsilon$, como queríamos probar. □

Observar que en esta demostración solamente se utilizaron las definiciones de convergencia y de sucesión de Cauchy, por lo que el resultado es válido en cualquier conjunto donde se puedan definir estas nociones.

Teorema 2.29. Si a_n es una sucesión de Cauchy en $\mathbb{R}$, entonces es convergente.

Demostración. Veamos primero que a_n es acotada. Para esto, tomemos un ε cualquiera, por ejemplo $\varepsilon = 1$. Entonces a partir de un n_0, tenemos en particular que $|a_n - a_{n_0}| < 1$, y por lo tanto $a_n \in E(a_{n_0}, 1)$, por lo que está acotada. Como está acotada, en virtud del Teorema 2.24, posee una subsucesión convergente, a la que llamaremos $a_{n_k} \to L$.

Veamos que toda la sucesión también tiende a L:
Sea $\varepsilon > 0$, y como a_n es de Cauchy, existe un $n_0 \in \mathbb{N}$ tal que $|a_m - a_n| < \varepsilon/2$ para todo $n, m \geqslant n_0$. Ahora, como $a_{n_k} \to L$, existe n_{k_0} tal que $|a_{n_k} - L| < \varepsilon/2$ para todo $n_k \geqslant n_{k_0}$. Pero podemos elegir este n_{k_0} de manera que además se cumpla $n_{k_0} \geqslant n_0$. Tomemos un $n \geqslant n_0$ y veamos que a_n está a menos de ε de L. En efecto, como n y n_{k_0} son mayores a n_0, por la condición de Cauchy, tenemos que $|a_n - a_{n_{k_0}}| < \varepsilon/2$. Entonces:

$$|a_n - L| = |a_n - a_{n_{k_0}} + a_{n_{k_0}} - L| \leqslant |a_n - a_{n_{k_0}}| + |a_{n_{k_0}} - L| \leqslant \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

□

En esta demostración sí hicimos uso del Axioma de completitud (¿dónde?), por lo que este resultado no es cierto para sucesiones en $\mathbb{Q}$ por ejemplo.

Ejercicio 2.30
Encontrar una sucesión de números racionales, que sea de Cauchy pero no convergente (a un número de $\mathbb{Q}$). Puede ser útil volver a mirar alguna sucesión estudiada en los ejemplos de este capítulo.

2.1.2. Sucesiones y continuidad

Dijimos que las sucesiones servían también para caracterizar objetos matemáticos, pero, salvo esta última discusión, no han aparecido en este papel. En este contexto, caracterizar quiere decir describir completamente una clase de objetos en función de las sucesiones. Es como una definición alternativa, una propiedad que podría perfectamente ser la definición. Por ejemplo, en el próximo resultado, que vamos a demostrar en un contexto más amplio en el Capítulo 5, damos una caracterización de las funciones continuas. Específicamente, las funciones continuas en un punto a son aquellas en las que, cuando nos acercamos por cualquier sucesión al punto a, las imágenes de la sucesión por f se acercan a $f(a)$.

Teorema 2.31. Una función $f : \mathbb{R} \to \mathbb{R}$ es continua en un punto a sii para toda sucesión $a_n \to a$, tenemos que $f(a_n) \to f(a)$.

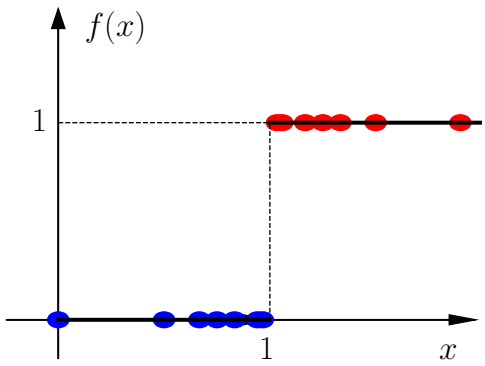


Figura 2.1: En rojo, algunos puntos de la forma $(a_n, f(a_n))$, y en azul algunos puntos de la forma $(b_n, f(b_n))$.

Ejemplo 2.32
Tomemos una función discontinua en un punto:

$$f(x) = \begin{cases} 1 & si\ x \geqslant 1 \\ 0 & si\ x < 0 \end{cases}$$

Para ver, mediante la caracterización por sucesiones, que la función no es continua en $x = 1$, tomemos dos sucesiones que tiendan a 1, pero una por la izquierda y otra por la derecha. Por ejemplo, $a_n = 1 + \frac{1}{n}$ y $b_n = 1 - \frac{1}{n}$.

Claramente $\lim a_n = \lim b_n = 1$, pero sin embargo $f(a_n) = 1$ para todo n, es decir, es una sucesión constante. Análogamente tenemos que $f(b_n) = 0$ para todo n, y por lo tanto tenemos que

$$0 = \lim_{n\to\infty} f(b_n) \neq \lim_{n\to\infty} f(a_n) = 1.$$

En el siguiente ejercicio utilizamos esta caracterización de las funciones continuas para demostrar una parte del Teorema de Weierstrass. Se pueden repetir argumentos similares para demostrar la segunda mitad.

Ejercicio 2.33

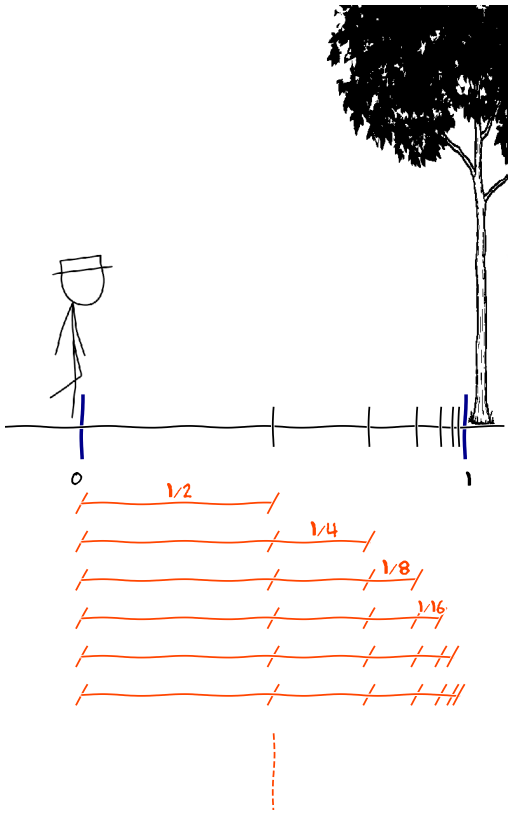
1. *Escribir la negación de acotación de una función.*
2. *Demostrar que si una función f no está acotada, se puede encontrar una sucesión a_n en el dominio de f tal que $f(a_n) \to \infty$.*
3. *Si ahora el dominio de la función es $[a,b]$, ¿qué se puede decir sobre la sucesión a_n construida en el item anterior?*
4. *Si ahora además la función es continua, estudiar qué sucede con las imágenes de alguna subsucesión conveniente, y concluir que la función tiene que ser acotada.*

Estudiaremos más la relación entre continuidad y sucesiones, en un marco más general, en el capítulo 5.

2.2. Series

Supongamos que queremos caminar hasta un árbol que se encuentra a un kilómetro de distancia desde donde estamos.

Entonces, para llegar hasta el árbol, primero debemos caminar medio kilómetro y llegar hasta la mitad del camino (ver figura[4]). Una vez allí, nos resta caminar $1/2km$. Ahora nuevamente, para caminar ese tramo, primero hay que caminar la mitad del mismo, es decir, $250m$, y una vez allí, nos restará caminar $125m$. Si repetimos este argumento sucesivamente, siempre nos quedará la mitad del camino por recorrer, y debemos recorrer infinitos tramos (de longitud cada vez menor), antes de llegar al árbol. Pero está claro que podemos llegar al árbol. Entonces, ¿cómo una suma de infinitos trayectos puede dar lugar a un camino que podemos recorrer?. Esta es una versión de la paradoja de Zenón, en cuya época todavía no se conocía el cálculo infinitesimal por supuesto.



Otra interpretación geométrica de esta serie surge de intentar pintar un cuadrado de lado 1, pintando cada vez un rectángulo de la mitad del área por pintar, como se ve en la figura 2.2.

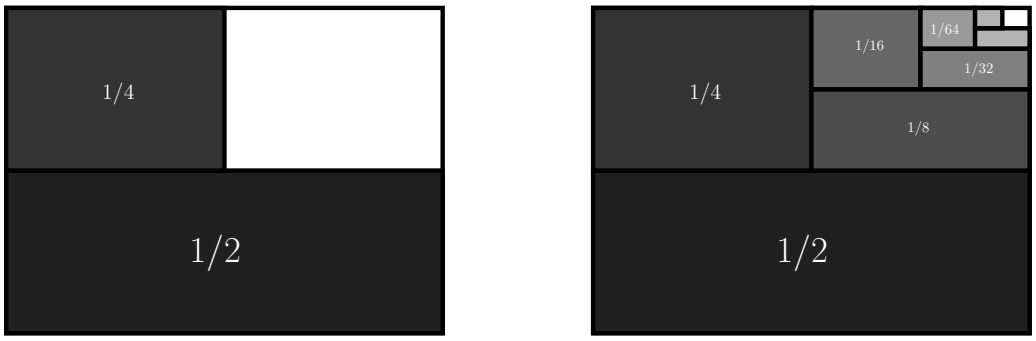
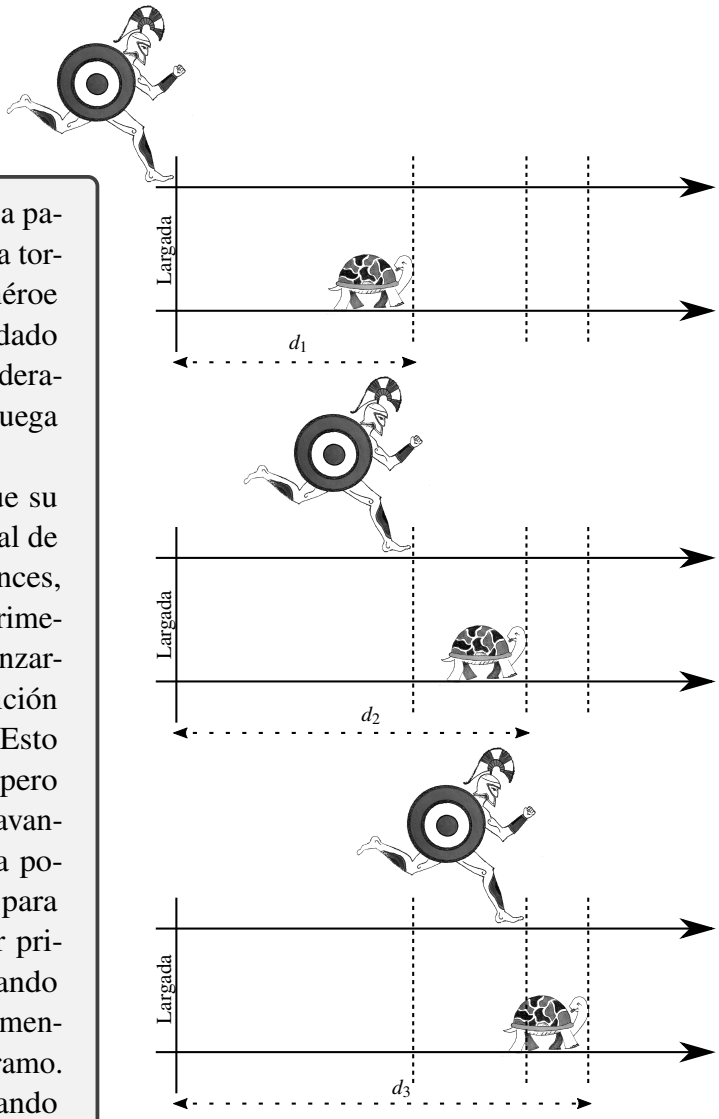


Figura 2.2: Cuadrado de lado 1 pintado en sucesivos rectángulos. A la izquierda, los dos primeros pasos. A la derecha, algunas iteraciones más tarde.

[4]Figura creada a partir de imágenes y estilo de la tira cómica xkcd.com.

Quizás la versión más conocida de la paradoja de Zenón es la de Aquiles y la tortuga. En este planteo, Aquiles, un héroe de la Guerra de Troya que era apodado "el de los pies ligeros" por ser considerado el más veloz entre los hombres, juega una carrera contra una tortuga.
Aquiles, sabiéndose más rápido que su contrincante, le da una ventaja inicial de algunos metros, digamos d_1. Entonces, para que Aquiles pase a la tortuga, primero la tiene que alcanzar. Y para alcanzarla, primero tiene que llegar a la posición d_1 desde donde partió la tortuga. Esto le llevará a Aquiles cierto tiempo, pero cuando llegue allí, la tortuga habrá avanzado un tramo y estará ahora en la posición d_2. Entonces, ahora Aquiles para alcanzar a la tortuga, deberá llegar primero hasta la posición d_2. Para cuando Aquiles logre llegar hasta allí, nuevamente la tortuga habrá avanzado algún tramo. Así, siempre Aquiles estará intentando alcanzar a la tortuga.



Volviendo a la versión del árbol, en definitiva, primero se camina una distancia de $\frac{1}{2}$, luego una de $\frac{1}{4}$, luego de $\frac{1}{8}$, y así sucesivamente. Siguiendo con los infinitos tramos, la distancia total recorrida es

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$$

Sabemos sumar una cantidad arbitraria (pero finita) de números. Pero, ¿qué quiere decir sumar infinitos números? ¿Qué propiedades tiene esa nueva suma que tenemos que definir? ¿Será asociativa? ¿conmutativa?. Adentrémonos en las series, y veamos que la suma infinita de arriba es uno, es decir, la distancia original al árbol.

La idea para definir una suma infinita será cortar la suma, mirar el acumulado de los primeros n sumandos, y hacer tender n a infinito:

Definición 2.34. Dada una sucesión real a_n, se llama *suma parcial* o *reducida enésima* a la sucesión $s_n = \sum_{i=1}^{n} a_i$, y se denomina *serie* de término general a_n a la suma infinita $\sum a_n$. Se dice que la serie converge, diverge, u oscila, cuando lo hace la sucesión s_n. Además, cuando la serie es convergente, al límite $S \in \mathbb{R}$ de s_n se lo denomina la suma de la serie, y se escribe $S = \sum a_n$. También utilizaremos la notación $\sum a_n < \infty$ para referirnos a que la serie es convergente.

Importante

Es importante no confundir la serie con el término general de la misma. Si a_n es una sucesión, la nueva sucesión s_n de sumas parciales básicamente acumula todos los términos de a_n hasta el índice n. Si la sucesión a_n tiene límite, la serie puede tener un comportamiento muy distinto.

El siguiente ejemplo puede ser muy básico para los que ya han trabajado con series, pero puede ser muy ilustrativo para aquellos que se enfrentan a las sumas infinitas y reducidad enésimas por primera vez. Consideremos la serie de término general $a_n = 1/n$, y pongamos en una tabla, en paralelo, cada término de la sucesión a_n (los primeros 5), y a su lado cada término de la reducida enésima s_n (el total acumulado hasta el momento):

n	a_n	s_n	s_n
1	1	1	1
2	$1/2$	$1+1/2=3/2$	1,5
3	$1/3$	$1+1/2+1/3=11/6$	$\approx 1{,}8333$
4	$1/4$	$1+1/2+1/3+1/4=25/12$	$\approx 2{,}0833$
5	$1/5$	$1+1/2+1/3+1/4+1/5=137/60$	$\approx 2{,}2833$

Observar, en particular, que a_n tiende a cero como ya sabíamos (la segunda columna), mientras que s_n es creciente. A priori no sabemos si s_n puede converger a un número finito, o si diverge a infinito.

Estudiemos ahora algunas series para las cuales sí podremos calcular el límite de la reducida enésima.

Ejemplos 2.35

1. *La serie $\sum_{n=0}^{\infty} q^n$ se denomina serie geométrica. Cuando $q = \frac{1}{2}$, tenemos la serie de la paradoja de Zenón (a menos del primer sumando). Si s_n es la reducida enésima, entonces $s_n = 1 + q + q^2 + q^3 + \cdots + q^n$. Ahora, multiplicando por q, tenemos: $qs_n = q + q^2 + q^3 + \cdots + q^n + q^{n+1}$, y restando:*

$$\begin{aligned} s_n &= 1 + \cancel{q} + \cancel{q^2} + \cancel{q^3} + \cdots + \cancel{q^n} \\ -qs_n &= \cancel{q} + \cancel{q^2} + \cancel{q^3} + \cdots + \cancel{q^n} + q^{n+1} \\ \hline s_n(1-q) &= 1 - q^{n+1} \end{aligned}$$

   Si $q \neq 1$, podemos despejar una expresión explícita para la reducida enésima: $s_n = \frac{1-q^{n+1}}{1-q}$. Ahora tenemos que calcular el límite de s_n. Cuando $|q| < 1$, tenemos que $\lim s_n = \frac{1}{1-q}$.

   Cuando $|q| > 1$, la serie no es convergente.

   ¿Qué pasa cuando $q = 1$ o $q = -1$?

2. *Consideremos ahora la serie $\sum_{n=1}^{\infty} (-1)^{n+1}$. La reducida enésima oscila entre 1 y 0, dependiendo de la paridad de n, por lo que la serie no es convergente.*

Proposición 2.36 (Condición necesaria de convergencia)**.** Si $\sum a_n$ es convergente, entonces $\lim a_n = 0$.

Demostración. Si s_n es la reducida enésima, entonces tenemos que $a_n = s_n - s_{n-1}$. Pero si $\sum a_n = S$, entonces $\lim a_n = \lim s_n - lim s_{n-1} = S - S = 0$. □

Es decir, para que una serie sea convergente, lo mínimo que tenemos que pedir es que su término general tienda a cero. Por lo tanto esto es lo primero que tenemos que chequear cuando estudiamos la convergencia de una serie.

Importante

Pero cuidado que es una condición necesaria pero no suficiente. Concluir la convergencia de la serie a partir de que $\lim a_n = 0$ es un error grave, como veremos en varios ejemplos.

Ejemplos 2.37

En algunos casos, cuando escribimos la reducida enésima, los términos se comienzan a cancelar entre sí, sobreviviendo solamente una cantidad finita de ellos, y por lo tanto haciendo posible el cálculo del límite de s_n. Reciben el nombre de series telescópicas, y aquí presentamos dos ejemplos.

1. $\sum \frac{1}{n(n+1)}$. *Observemos primero que podemos re-escribir* $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$. *Entonces la reducida enésima es:*

$$s_n = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right).$$

De donde $s_n = 1 - \frac{1}{n+1}$, *y por lo tanto la serie converge a uno.*

2. $\sum \log\left(1 + \frac{1}{n}\right) = \sum \log\left(\frac{n+1}{n}\right) = \sum (\log(n+1) - \log(n))$. *De donde la reducida enésima es*

$$s_n = (\cancel{\log(2)} - \log(1)) + (\cancel{\log(3)} - \cancel{\log(2)}) + (\cancel{\log(4)} - \cancel{\log(3)}) + \cdots + (\log(n+1) - \cancel{\log(n)}).$$

De donde $s_n = \log(n+1)$, *y entonces podemos calcular* $\lim s_n = \infty$, *por lo que la serie es divergente.*

En general, es muy difícil hallar exactamente la suma de una serie (es decir, el valor al que converge la sucesión de sumas parciales), y solamente en contados casos se puede calcular. En este sentido, los ejemplos anteriores no son representativos de la mayoría de las series, aunque de todas maneras resultarán muy útiles. Por lo tanto, muchas veces nos conformamos con poder decidir si una serie es convergente o no, y eventualmente acotar el valor al que converge. En lo que sigue, veremos herramientas para clasificar series, basándonos explícita o implícitamente en otras series conocidas.

2.2.1. Series de términos positivos

Comenzaremos con los resultados más fuertes, que son para series de términos positivos, es decir, cuando el término general es $a_n \geqslant 0$. Observar que en este caso, la sucesión s_n de sumas parciales es monótona creciente, por lo tanto la serie $\sum a_n$ puede ser convergente o divergente, pero no puede oscilar.

Proposición 2.38 (Criterio de comparación)**.** Sean $\sum a_n$ y $\sum b_n$ series de términos positivos, tales que $a_n \leqslant b_n$ para todo $n > n_0$. Entonces:

- Si $\sum b_n$ converge $\Rightarrow \sum a_n$ converge.
- Si $\sum a_n$ diverge $\Rightarrow \sum b_n$ diverge.

Demostración. Llamemos A_n y B_n a las reducidas enésimas: $A_n = a_1 + a_2 + \cdots + a_n$, y $B_n = b_1 + b_2 + \cdots + b_n$. Entonces $A_n - A_{n_0} \leqslant B_n - B_{n_0}$ a partir de n_0. Veamos el primer caso: si $\sum b_n$ converge, en particular B_n está acotada. Por lo tanto A_n está acotada, y por ser monótona creciente, es convergente. Si por otro lado $\sum a_n$ es divergente, entonces A_n es no acotada, y por lo tanto también lo será B_n. Como además B_n es creciente, entonces necesariamente es divergente. □

Observar que en esta demostración (y esto es general en las series), los primeros n_0 términos no importan. Los que importan son los "últimos" términos de la serie. Si tomamos por ejemplo la serie $\sum \frac{1}{2^n}$, y cambiamos el primer millón de términos por 2000, la nueva serie igual tendrá suma finita. La convergencia de la serie se juega en el infinito.

Los siguientes resultados van en el mismo sentido, especialmente el próximo, pero antes veamos algunos ejemplos del criterio de comparación.

Ejemplos 2.39

1. *Estudiemos la serie* $\sum \frac{1}{n}$*, que se denomina serie armónica. Sabemos (ver por ejemplo el ejercicio 97 en el capítulo de Integrales de las notas del curso [8]) que* $\log(1+x) \leqslant x$*, de donde tenemos* $\log\left(1+\frac{1}{n}\right) \leqslant \frac{1}{n}$*, y como ambas sucesiones son no negativas, podemos usar el criterio de comparación. Como la serie de término general* $\log\left(1+\frac{1}{n}\right)$ *diverge (ver ejemplo 2.37), concluimos que* $\sum \frac{1}{n}$ *diverge. Observar que el término general* $\frac{1}{n}$ *tiende a cero, pero la serie no es convergente.*

2. *Si* $\alpha < 1$*, entonces* $\frac{1}{n^\alpha} > \frac{1}{n}$*, y como* $\sum \frac{1}{n}$ *diverge, por comparación, también lo hacen las series* $\sum \frac{1}{n^\alpha}$ *con* $\alpha < 1$*.*

Veamos ahora que si tenemos dos sucesiones que se comportan igual en el infinito, entonces las series asociadas son de la misma clase. Esto quiere decir que ambas son convergentes, o ambas son divergentes (como estamos trabajando con series de términos positivos, no hay otra posibilidad, pues no puede oscilar).

Proposición 2.40 (Criterio de equivalentes)**.** Sean $\sum a_n$ y $\sum b_n$ dos series de términos positivos.

- Si $\lim \frac{a_n}{b_n} = L > 0$, entonces las dos series son de la misma clase.
- Si $\lim \frac{a_n}{b_n} = 0$, entonces si $\sum b_n$ converge, $\sum a_n$ también lo hace.

Demostración. Supongamos primero que $\lim \frac{a_n}{b_n} = L > 0$. Entonces (tomando $\varepsilon = L/2$ en la definición) a partir de un n_0, tenemos $\frac{L}{2} \leqslant \frac{a_n}{b_n} \leqslant \frac{3L}{2}$, y por lo tanto podemos acotar: $b_n\frac{L}{2} \leqslant a_n \leqslant b_n\frac{3L}{2}$. Ahora podemos utilizar el criterio de comparación dos veces, con las dos desigualdades. Específicamente:
Si $\sum b_n < \infty$, como $a_n \leqslant b_n\frac{3L}{2}$, el criterio de comparación afirma que $\sum a_n < \infty$.
Si por el contrario $\sum b_n$ diverge, como $b_n\frac{L}{2} \leqslant a_n$, el criterio de comparación afirma que $\sum a_n$ diverge.

El caso $\lim \frac{a_n}{b_n} = 0$ es similar, y queda como ejercicio. □

Con este resultado entonces, para estudiar la convergencia de una serie, basta estudiar una serie cuyo término general sea equivalente. Así, cualquier serie cuyo término a_n sea un cociente de polinomios, por ejemplo será muy sencillo de clasificar.

Ejemplos 2.41

1. *Como* $\frac{1}{n^2} \sim \frac{1}{n(n+1)}$, *y la serie* $\sum \frac{1}{n(n+1)} < \infty$ *(ver ejemplo 2.37), tenemos que la serie* $\sum \frac{1}{n^2}$ *es convergente.*

2. *Si* $\alpha > 2$, *entonces* $\frac{1}{n^\alpha} \leqslant \frac{1}{n^2}$, *y como vimos que* $\sum \frac{1}{n^2} < \infty$, *tenemos que la serie* $\sum \frac{1}{n^\alpha}$ *es convergente para* $\alpha > 2$. *De esta manera, tenemos clasificadas las series de la forma* $\sum \frac{1}{n^\alpha}$, *salvo para valores de* $\alpha \in (1,2)$. *Volveremos a esto más adelante.*

3. *La serie* $\sum \sin\left(\frac{1}{n}\right)$ *es divergente, pues el término general es equivalente a* $\frac{1}{n}$.

4. $\sum \frac{1}{\sqrt{n(n+2)}}$ *es divergente, pues el término general es equivalente a* $\frac{1}{n}$.

5. $\sum \frac{1}{\sqrt{n(n+1)(n+2)(n+3)}}$ *es convergente, pues el término general es equivalente a* $\frac{1}{n^2}$

Proposición 2.42 (Criterio del cociente o criterio de D'Alambert)**.** Sea $\sum a_n$ una serie de términos positivos, tal que existe $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = L$. Entonces:

- Si $L < 1 \Rightarrow \sum a_n < \infty$.
- Si $L > 1 \Rightarrow \sum a_n$ diverge.

Demostración. Si $L < 1$, entonces a partir de cierto n_0, tenemos que $\frac{a_{n+1}}{a_n} \leqslant k < 1$,[5] y por lo tanto $a_{n+1} \leqslant k a_n$. Podemos utilizar la desigualdad recursivamente para ir desde n hasta n_0:

$$a_n \leqslant k a_{n-1} \leqslant k^2 a_{n-2} \leqslant \cdots \leqslant k^{n-n_0} a_{n_0} = \frac{a_{n_0}}{k^{n_0}} k^n.$$

Luego, como $0 \leqslant k < 1$, entonces la serie geométrica $\sum k^n$ converge, y por comparación también lo hace $\sum a_n$.

Si $L > 1$, entonces a partir de cierto $n_0 \in \mathbb{N}$, tenemos $a_{n+1} \geqslant a_n$.[6] En particular, $a_n \geqslant a_{n_0} > 0$ para todo $n > n_0$, por lo que la sucesión no puede tender a cero, y entonces $\sum a_n$ diverge. □

[5] Podemos tomar un ε tal que $L + \varepsilon < 1$, y por lo tanto a partir de cierto n_0 la sucesión $\frac{a_{n+1}}{a_n}$ estará entre $L - \varepsilon$ y $L + \varepsilon$. Llamando $k = L + \varepsilon$, tenemos lo buscado.

[6] Nuevamente, podemos tomar ε tal que $L - \varepsilon > 1$, y entonces a partir de cierto n_0 tenemos que $\frac{a_{n+1}}{a_n} > L - \varepsilon > 1$.

Observar que cuando el límite es 1, el criterio no decide: la serie podría ser convergente o divergente, y tendremos que usar otro método para clasificarla.

Ejemplos 2.43

1. *Estudiemos la serie* $\sum \frac{n!}{n^n}$. *El cociente* $\frac{a_{n+1}}{a_n}$ *es:*

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n = \frac{1}{\left(1+\frac{1}{n}\right)^n}.$$

   Del ejemplo 2.15, concluimos que $\lim \frac{a_{n+1}}{a_n} = \frac{1}{e}$, *y por lo tanto la serie es convergente.*

2. *Para la serie* $\sum \frac{2^n}{n!}$, *tenemos:*

$$\frac{a_{n+1}}{a_n} = \frac{2^{n+1}}{(n+1)!} \cdot \frac{n!}{2^n} = \frac{2}{(n+1)}.$$

   De donde $\frac{a_{n+1}}{a_n} \to 0$, *y por lo tanto la serie es convergente.*

3. *Ya hemos clasificado las series* $\sum \frac{1}{n}$ *y* $\sum \frac{1}{n^2}$. *Veamos qué dice el criterio del cociente en estos casos. Para* $a_n = \frac{1}{n}$, *tenemos:*

$$\lim \frac{a_{n+1}}{a_n} = \lim \frac{n}{n+1} = 1.$$

   Y para $a_n = \frac{1}{n^2}$*:*

$$\lim \frac{a_{n+1}}{a_n} = \lim \frac{n^2}{(n+1)^2} = 1.$$

   Es decir, el criterio no decide en ninguno de los dos casos (pues tenemos $L = 1$*), pero sabemos (porque las clasificamos mediante otros mecanismos) que una de las series que es convergente, y la otra divergente.*

Veamos ahora un criterio similar, también para series de términos positivos.

Proposición 2.44 (Criterio de la raíz enésima o criterio de Cauchy)**.** Sea $\sum a_n$ una serie de términos positivos, tal que existe $\lim_{n\to\infty} \sqrt[n]{a_n} = L$. Entonces:

- Si $L < 1 \Rightarrow \sum a_n < \infty$.
- Si $L > 1 \Rightarrow \sum a_n$ diverge.

Demostración. Si $L < 1$, entonces a partir de cierto $n_0 \in \mathbb{N}$, tenemos que $\sqrt[n]{a_n} \leqslant k < 1$,[7] y por lo tanto $a_n \leqslant k^n$. Como $k < 1$, la serie geométrica $\sum k^n$ es convergente, y por comparación, también lo es $\sum a_n$.

Si $L > 1$, entonces $\sqrt[n]{a_n} \geqslant 1$ a partir de cierto n_0, lo que implica que $a_n \geqslant 1$, y entonces la serie $\sum a_n$ es divergente porque el término general no tiende a cero. □

[7] Aquí utilizamos el mismo argumento que en la prueba de 2.42 para encontrar un k que cumpla lo enunciado.

Se puede ver que el criterio de Cauchy es más fuerte que el criterio de D'Alambert, en el sentido que si una serie se puede clasificar mediante el criterio del cociente, entonces también se puede clasificar por el criterio de la raíz enésima. Sin embargo, en general el criterio del cociente es más fácil de chequear.

Ejemplos 2.45

1. $\sum \left(\frac{\log n}{n}\right)^n$ *es convergente pues* $\lim \sqrt[n]{a_n} = \lim \sqrt[n]{\left(\frac{\log n}{n}\right)^n} = \lim \frac{\log n}{n} = 0.$

2. $\sum \frac{n^5}{2^n}$ *es convergente pues* $\lim \sqrt[n]{a_n} = \lim \sqrt[n]{\frac{n^5}{2^n}} = \frac{1}{2}.$

Y esta serie ¿converge?

Antes de pasar a series alternadas, aprovechemos esta oportunidad para mencionar una serie particular, que lleva el nombre de *Flint Hill*, como excusa para ejemplificar la cantidad de preguntas abiertas que hay en matemática, la cantidad de cosas que no se saben.
La *Flint Hill Series* es

$$\sum_{n=1}^{\infty} \frac{1}{n^3 \sin^2(n)},$$

y no se sabe si es convergente o no.
Como sucede en no pocas ocasiones, es un problema fácil de enunciar (al punto que podría pasar desapercibida en un práctico de cálculo), pero cuya respuesta es aún desconocida.

2.2.2. Series alternadas

Vamos a estudiar ahora series de términos que no necesariamente son de signo constante.

Definición 2.46. Decimos que una serie $\sum a_n$ es *absolutamente convergente* sii $\sum |a_n|$ es convergente.

Por ejemplo, si tomamos $-1 < a < 1$, entonces la serie geométrica $\sum a^n$ es absolutamente convergente, pues $|a^n| = |a|^n$, y como $|a| < 1$, entonces $\sum |a|^n < \infty$.

Intuitivamente, si una serie es absolutamente convergente, debería ser convergente también $\sum a_n$, pues sumar el valor absoluto de los términos es el "peor caso" de alguna manera. Si hay términos positivos y negativos, algunos se podrán compensar con otros. Si todos los términos tienen el mismo signo, están todos aportando hacia el mismo lado. Formalicemos este resultado:

Teorema 2.47. Toda serie absolutamente convergente es convergente.

Demostración. Sea $\sum a_n$ la serie absolutamente convergente, y separemos los términos a_n en los positivos y negativos, y formemos dos nuevas sucesiones:

$$a_n^+ = \begin{cases} a_n & \text{si } a_n \geqslant 0 \\ 0 & \text{si } a_n < 0 \end{cases} \qquad a_n^- = \begin{cases} 0 & \text{si } a_n \geqslant 0 \\ -a_n & \text{si } a_n < 0 \end{cases}$$

Es decir, en a_n^+ ponemos los términos positivos de a_n, y los negativos los ponemos a cero, mientras que en a_n^- ponemos los términos positivos a cero, y copiamos los negativos pero cambiándole el signo. De esta manera, las dos nuevas sucesiones son de términos positivos.

Observar que tenemos: $a_n = a_n^+ - a_n^-$, y $|a_n| = a_n^+ + a_n^-$.

Como $\sum |a_n|$ converge, y tenemos $0 \leqslant a_n^+ \leqslant |a_n|$, por el criterio de comparación, $\sum a_n^+$ converge. De la misma manera se obtiene que $\sum a_n^-$ converge, y por lo tanto también converge $\sum (a_n^+ - a_n^-) = \sum a_n$. □

El recíproco de este resultado no es cierto. Vamos a ver luego que hay series que son convergentes, pero que la serie de valores absolutos de a_n no lo es. A estas series que son convergentes pero no absolutamente convergentes, se les denomina *condicionalmente convergentes*.

Cuando tenemos una serie de términos positivos, la convergencia absoluta no agrega nada. Por lo tanto este resultado lo vamos a usar cuando tenemos una serie con términos que cambian de signo, pero podemos clasifica la serie de valores absolutos, utilizando algunas de las herramientas que vimos para series de términos positivos.

Ejemplo 2.48
$\sum \frac{\sin(n)}{n^2}$ *no es de términos positivos (pues* $\sin(n)$ *cambia de signo), pero cuando miramos la serie de los valores absolutos, podemos acotarla por una convergente:* $\frac{|\sin(n)|}{n^2} \leqslant \frac{1}{n^2}$. *Por lo tanto,* $\sum \frac{|\sin(n)|}{n^2}$ *es convergente por comparación, y* $\sum \frac{\sin(n)}{n^2}$ *es convergente por el Teorema 2.47.*

El siguiente resultado, que no demostraremos aquí, es el único criterio que veremos para clasificar series que no son de signo constante (con excepción de la condición necesaria de convergencia).

Proposición 2.49 (Criterio de Leibnitz)**.** Si a_n es una sucesión monótona decreciente que tiende a cero, entonces la serie alternada $\sum(-1)^n a_n$ es convergente.

Observar que con este criterio tenemos clasificadas muchísimas series de término general $(-1)^n a_n$, porque si esta sucesión a_n tiende a cero de forma monótona, podemos aplicar el criterio de Leibnitz para concluir su convergencia. Y si no tiende a cero, entonces no cumple la condición necesaria de convergencia (pues $(-1)^n a_n \nrightarrow 0$), y por lo tanto es divergente.

Ejemplos 2.50

1. $\sum(-1)^n \frac{1}{n}$ *es convergente por el criterio de Leibnitz, pero no es absolutamente convergente, pues* $\sum\left|(-1)^n \frac{1}{n}\right| = \sum\frac{1}{n}$*, que es divergente como ya vimos.*
2. $\sum(-1)^n \log\left(1+\frac{1}{n}\right)$ *es convergente por el criterio de Leibnitz.*

Extra: Reordenación de series

Cuando generalizamos conceptos, como en este caso la suma de una cantidad finita a una cantidad infinita de términos, se pueden perder algunas de las propiedades que teníamos. En este caso, veremos qué sucede con la conmutatividad de la suma.
Por supuesto que la suma de una cantidad finita de términos es conmutativa. ¿Qué sucede con las series?
Si la serie es *absolutamente convergente*, entonces cualquier reordenación de lol términos da el mismo resultado. Es decir, en este caso conservamos la conmutatividad: no importa el orden de los sumandos.
Sin embargo, si la serie es *condicionalmente convergente*, el resultado sí depende del orden de los términos. Más aún, el siguiente resultado nos dice que podemos reordenar los términos de la serie para que sea convergente a cualquier número real que elijamos.

Teorema 2.51 (Riemann)**.** Si $\sum a_n$ es una serie condicionalmente convergente, y L es un número real cualquiera, entonces existe una reordenación de $\sum a_n$ que converge a L.

La demostración de este teorema no es difícil. Básicamente consiste en tomar en orden los términos positivos, hasta que su suma sobrepase el valor L. Luego, se continúa con los términos negativos (también en orden) hasta que la suma total (considerando los positivos ya agregados) sea menor que L. Esto se puede hacer ya que tanto la serie de términos positivos como la de términos negativos son divergentes. El argumento se repite: volvemos a tomar términos positivos (retomando donde habíamos quedado) hasta sobrepasar L. Luego términos negativos, y así sucesivamente. La serie reordenada converge a L por construcción.

Capítulo 3

Integrales impropias

En la última sección formalizamos la idea de sumar infinitos términos, y en este capítulo desarrollaremos una idea similar, que en algún sentido es una versión continua de las series, pues en lugar de sumar vamos a integrar.

Cuando fue definida la integral, teníamos que tener en cuenta dos aspectos: que el intervalo de integración fuera acotado y que la función fuera acotada en ese dominio. En lo que sigue vamos a levantar estas consideraciones, dando lugar a las denominadas integrales impropias (de primera y segunda especie, respectivamente).

La integral impropia aparece en muchas definiciones, como por ejemplo en la transformada de Fourier, que es muy utilizada tanto en ingeniería como en matemática pura. En este curso nos ocuparemos solamente de nociones básicas y ejemplos, para que en otros cursos utilicen las integrales impropias como herramienta.

3.1. Integrales impropias de primera especie

Empecemos primero con las integrales impropias de primera especie, esto es, queremos darle sentido a $\int_a^\infty f(t)dt$. El procedimiento será idéntico al que llevamos a cabo en el capítulo anterior: integraremos hasta un valor finito b, y luego tomaremos límite cuando b tiende a infinito.

Definición 3.1. Sea $f : [a,\infty) \to \mathbb{R}$ continua, y llamemos $F(x) = \int_a^x f(t)dt$. Entonces si existe $\lim_{x\to\infty} F(x)$ y es finito, decimos que la integral impropia $\int_a^\infty f(t)dt$ es convergente a ese valor. Si el límite es infinito, decimos que la integral impropia diverge, y si el límite no existe, decimos que la integral impropia oscila.

Observar que la función $F(x)$, que acumula la integral desde a hasta x, juega el papel

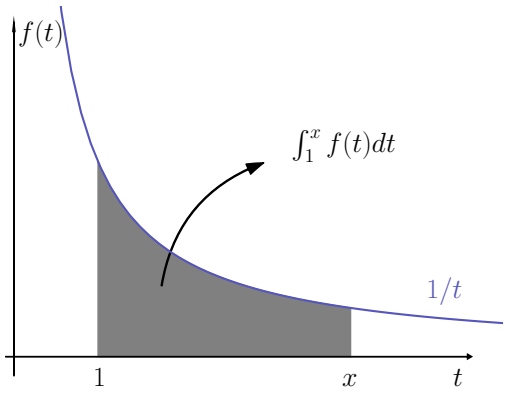
que jugaba la reducida enésima en las series. Tanto es así, que varios de los resultados para impropias se demuestran de forma idéntica al capítulo pasado, sustituyendo la reducida enésima por $F(x)$.

Si bien en la definición pedimos que f sea continua, en realidad es suficiente con que sea integrable en cualquier subconjunto acotado del dominio, ya que eso es lo que estamos haciendo: integrando en conjuntos acotados (que es lo que ya sabíamos hacer), y luego hacer un límite.

Ejemplos 3.2

1. *El primer caso que estudiaremos será uno de los más útiles, se trata de la impropia* $\int_1^\infty \frac{1}{x^\alpha}dx$. *Calculemos entonces la primitiva*

$$F(x) = \int_1^x \frac{1}{t^\alpha}dt = \begin{cases} \frac{x^{1-\alpha}-1}{1-\alpha} & si\ \alpha \neq 1 \\ \log(x) & si\ \alpha = 1 \end{cases}.$$



Entonces si $\alpha > 1$, *tenemos*

$$\lim_{x\to+\infty} F(x) = \frac{1}{\alpha - 1},$$

y por lo tanto la integral impropia es convergente $\int_1^\infty \frac{1}{x^\alpha}dx = \frac{1}{\alpha-1}$.

Si $\alpha \leqslant 1$, *la integral impropia es divergente.*

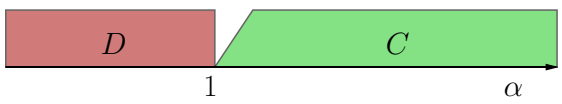


Figura 3.1: Cuando $\alpha \leqslant 1$, la impropia $\int_1^{+\infty} \frac{1}{x^\alpha}dx$ es divergente, y cuando $\alpha > 1$ es convergente.

Observar que tenemos el mismo comportamiento que en la serie $\sum \frac{1}{n^\alpha}$, *para los casos de* α *que pudimos clasificar en el capítulo anterior. Este ejemplo, junto con un resultado*

que relaciona las series con las integrales impropias, nos ayudará a clasificar la serie para los valores de α *en* $(1,2)$.

2. *Estudiemos la impropia* $\int_0^\infty \frac{1}{1+x^2}dx$. *Como* $F(x) = \int_0^x \frac{1}{1+t^2}dt = \arctan(x)$, *entonces tenemos:*
$$\int_0^\infty \frac{1}{1+x^2}dx = \lim_{x\to+\infty} \arctan(x) = \frac{\pi}{2}.$$

3. *La integral impropia* $\int_0^\infty \cos(x)dx$ *oscila, porque* $F(x) = \int_0^x \cos(t)dt = \sin(x)$ *no tiene límite cuando* $x \to \infty$.

De la linealidad de la integral y del límite, se desprende el siguiente resultado de linealidad para impropias:

Proposición 3.3. Si $\int_a^\infty f(t)dt$ y $\int_a^\infty g(t)dt$ convergen, entonces $\int_a^\infty (\alpha f(t) + \beta g(t))dt$ converge, y vale $\alpha \int_a^\infty f(t)dt + \beta \int_a^\infty g(t)dt$.

En el caso de las series teníamos que si $\sum a_n < \infty$, entonces necesariamente el término general a_n tiende a cero. Cabe preguntarse si tenemos un resultado similar para el caso de integrales impropias. Es decir, ¿seremos capaces de construir una función que no tienda a cero, pero cuya integral impropia sea convergente? El siguiente ejemplo muestra que con las funciones tenemos más libertad que con las sucesiones.

Ejemplo 3.4
Tomemos la función $f : [0,\infty) \to \mathbb{R}$ *definida como*[1]*:*

$$f(x) = \begin{cases} 1 & \text{si } x \in [n, n+\frac{1}{2^n}], \text{ con } n \in \mathbb{N} \\ 0 & \text{en otro caso.} \end{cases}.$$

Es decir, son escalones de altura constante, que empiezan en cada natural, y tienen un ancho cada vez menor (ver figura). Entonces si $F(x) = \int_0^x f(t)dt$, *es claro que* $F(n)$ *es la suma de las áreas de los primeros n escalones:* $F(n) = \sum_{k=0}^{n-1} \frac{1}{2^k}$, *que forma una serie geométrica que converge a* 2.

La integral impropia $\int_0^\infty f(t)dt$ *es convergente entonces, aunque la función* $f(x)$ *no tienda a cero cuando x tiende a infinito. Más adelante veremos otros ejemplos con un comportamiento similar.*

En este ejemplo, la función f no tiene límite en infinito. Sin embargo, si agregamos como hipótesis que el límite exista, entonces sí tenemos una condición similar a la que obtuvimos en el capítulo anterior.

[1] Observar que esta función no es continua, por lo que estamos haciendo uso de la definición más general comentada más arriba.

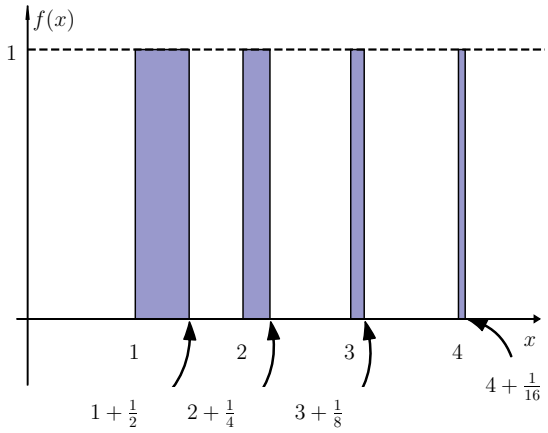


Figura 3.2: Función cuya integral impropia es convergente (es la suma de las áreas de los rectángulos, que resulta en una serie geométrica), pero que no tiene límite cuando x tiende a infinito.

Ejercicio 3.5
Sea f tal que $\int_a^\infty f(t)dt$ converge, y existe $\lim_{x\to\infty} f(x) = L$. Demostrar que $L = 0$.

A continuación vamos a ver resultados análogos a los criterios de comparación y de equivalentes vistos para series, que nos permitirán clasificar una gran cantidad de integrales impropias de signo constante. Las demostraciones son muy similares a los resultados correspondientes para series, y por lo tanto quedan como ejercicio.

Proposición 3.6. Sean f y g funciones continuas tales que $0 \leqslant f(t) \leqslant g(t)$ para todo $t > a$. Entonces:

- Si $\int_a^\infty g(t)dt$ converge, entonces $\int_a^\infty f(t)dt$ converge.
- Si $\int_a^\infty f(t)dt$ diverge, entonces $\int_a^\infty g(t)dt$ diverge.

Observar que solo nos importa que g sea mayor que f a partir de un cierto momento a. Nuevamente, lo que pasa "al principio" no nos importa, la convergencia se juega en el infinito.

Ejemplos 3.7

1. *A pesar de que ya calculamos el valor de $\int_0^\infty \frac{1}{x^2+1}dx$, podemos clasificarla observando que $\frac{1}{1+x^2} \leqslant \frac{1}{x^2}$, y como $\int_1^\infty \frac{1}{x^2}dx$ converge, también lo hace $\int_0^\infty \frac{1}{x^2+1}dx$. Observar que las integrales impropias comparadas no comienzan en el mismo punto. Sin embargo podemos utilizar el resultado (¿por qué?).*
2. *$\int_2^\infty \frac{1}{\log(x)}dx$ diverge, pues $\log(x) < x$ a partir de cierto punto. Entonces $\frac{1}{\log(x)} > \frac{1}{x}$, y como $\int_1^\infty \frac{1}{x}dx$ diverge, también lo hace $\int_2^\infty \frac{1}{\log(x)}dx$.*

3. *Clasifiquemos $\int_0^\infty e^{-x}x^k dx$. Como a partir de algún x_0 se cumple que $e^x > x^{k+n}$ para cualquier n, podemos elegir en particular $n = 2$, obteniendo que $e^{-x}x^k < \frac{1}{x^2}$. Como $\int_1^\infty \frac{1}{x^2}dx$ converge, también lo hace $\int_0^\infty e^{-x}x^k dx$, para cualquier $k \in \mathbb{N}$.*

   Observar que podríamos haber elegido $n = 3$ o cualquier otro mayor que dos. Aunque la desigualdad sea cierta también para $n = 1$ por ejemplo, esto no nos sirve porque llegaríamos a que la impropia que queremos clasificar es más chica que una que diverge, lo cual no nos dice nada.

Proposición 3.8. Sean f y g funciones con $f(t) \geqslant 0, g(t) \geqslant 0$ para todo t, y $\lim_{x\to\infty} \frac{f(x)}{g(x)} = L > 0$. Entonces $\int_a^\infty f(t)dt$ y $\int_a^\infty g(t)dt$ son de la misma clase.

Es decir, para clasificar una integral impropia de primera especie, basta estudiar el comportamiento de la función en el infinito.

Ejemplo 3.9

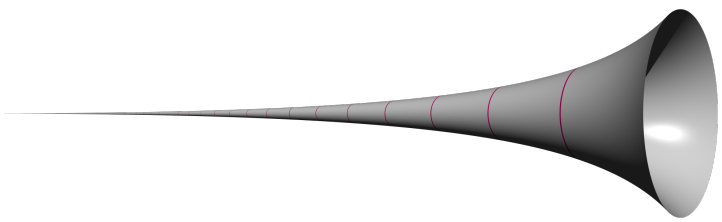
1. *Clasifiquemos $\int_0^\infty \frac{x}{\sqrt{x^4+1}}dx$. Cuando x tiende a infinito, el denominador es equivalente a x^2, por lo que el integrando es equivalente a $1/x$, y entonces la impropia es divergente.*

2. *Clasifiquemos $\int_0^\infty \frac{\sqrt{x}}{x^2+1}dx$. Nuevamente, cuando x tiende a infinito el denominador es equivalente a x^2. Luego, el integrando es equivalente a $\frac{1}{x^{3/2}}$, que es convergente.*

3. *Clasifiquemos $\int_1^\infty (e^{1/x} - 1 - 1/x)dx$. Para esto, queremos estudiar el comportamiento del integrando en el infinito. En este caso, cuando x tiende a infinito, como $1/x$ tiende a cero, podemos utilizar el desarrollo de Taylor de la exponencial en el origen. Entonces el integrando es equivalente a $\frac{1}{2x^2}$, y por lo tanto la integral impropia es convergente.*

Ejemplo 3.10

Las sumas o integrales infinitas, o muchas veces la noción del infinito en sí, están llenas de elementos que resultan poco intuitivos. La paradoja de Zenón es un ejemplo de eso, y la comunidad matemática tardó siglos en dar una explicación formal y convincente. También vimos el caso de la no conmutatividad de la suma en series condicionalmente convergentes. Otro ejemplo es la denominada trompeta de Torricelli, que consiste en tomar el sólido de revolución que se genera al girar el gráfico de la función $1/x$ en torno al eje x, como muestra la figura.



Se puede ver que, cuando se toma un intervalo $[1,a]$ y se realiza el procedimiento indicado, el volumen del sólido resultante se puede calcular como

$$V = \pi \int_1^a f(x)^2 dx.$$

Y la superficie del mismo, es:

$$A = 2\pi \int_1^a f(x)\sqrt{1+(f'(x))^2}dx.$$

En el caso de la trompeta de Torricelli, estas expresiones quedan:

$$V = \pi \int_1^a \frac{1}{x^2}dx, \qquad A = 2\pi \int_1^a \frac{\sqrt{1+\frac{1}{x^4}}}{x}dx.$$

Lo sorprendente de este ejemplo viene cuando consideramos la trompeta infinita. Entonces debemos calcular las integrales impropias que resultan de las expresiones anteriores, cuando a tiende a infinto. Estas son:

$$V_\infty = \pi \int_1^{+\infty} \frac{1}{x^2}dx, \qquad A_\infty = 2\pi \int_1^{+\infty} \frac{\sqrt{1+\frac{1}{x^4}}}{x}dx.$$

La primera de ellas ya la tenemos calculada, y es convergente: $V_\infty = \pi$. La segunda, sin embargo, es divergente (se puede ver que es equivalente a la impropia de $\frac{1}{x}$, o compararla con la misma). Tenemos por lo tanto una trompeta con volumen finito, pero cuya superficie es infinita.

La clara similitud entre las series y las integrales impropias de primera especie, que se ve en los resultados y en los ejemplos, la vamos a formalizar en el siguiente teorema, que es además una herramienta muy importante para clasificar.

Teorema 3.11 (Criterio Serie-Integral)**.** Sea $f : [k,+\infty) \to \mathbb{R}$ monótona decreciente, y $f(x) \geqslant 0$ para todo x. Entonces $\sum_{n=k}^{+\infty} f(n)$ y $\int_k^{+\infty} f(x)dx$ son de la misma clase.

Demostración. Supongamos que k es entero (o tomemos el primer entero mayor que k).

Como f es decreciente, en cada intervalo de la forma $[n,n+1]$ tenemos que

$$f(n+1) \leqslant f(t) \leqslant f(n), \forall t \in [n,n+1]$$

Si integramos en ese intervalo, resulta

$$\int_n^{n+1} f(n+1)dt \leqslant \int_n^{n+1} f(t)dt \leqslant \int_n^{n+1} f(n)dt$$

y como $f(n)$ y $f(n+1)$ son constantes,

$$f(n+1) \leqslant \int_n^{n+1} f(t)dt \leqslant f(n).$$

Ahora sumando desde k hasta n, y sustituyendo $f(i)$ por a_i, tenemos

$$\sum_{i=k+1}^{n+1} a_i \overset{(i)}{\leqslant} \int_k^{n+1} f(t)dt \overset{(ii)}{\leqslant} \sum_{i=k}^{n} a_i.$$

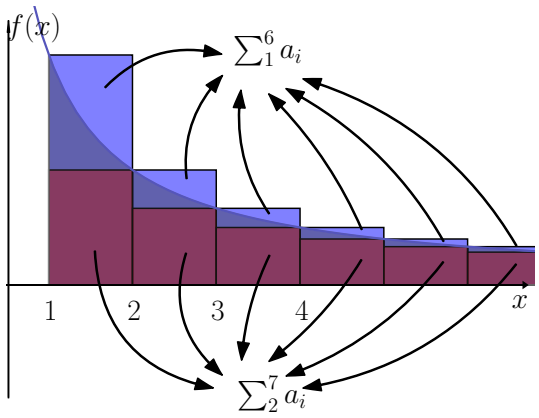


Figura 3.3: Interpretación geométrica de la desigualdad anterior.

Ahora, llamemos $F(x) = \int_k^{+\infty} f(x)dx$. Observar que F es creciente, pues f es positiva. Por lo tanto el límite de $F(x)$ cuando x tiende a infinito puede ser finito (en caso que F esté acotada) o infinito (en caso que F no esté acotada). Entonces:

- Si $\sum_k^{+\infty} f(n)$ es convergente, esto implica que $F(x)$ esté acotada por el valor de la serie (por la desigualdad (ii)), por lo que $\lim_{x\to\infty} F(x)$ es finito, y entonces la integral impropia es convergente.
- Si por el contrario $\sum_k^{+\infty} f(n)$ es divergente, entonces tenemos que $F(x)$ es no acotada (por la desigualdad (i)), de donde $\lim_{x\to\infty} F(x) = +\infty$, y entonces la integral impropia es divergente. □

En la mayoría de los casos es más fácil hallar una primitiva que calcular la reducida enésima, o incluso que clasificar la serie mediante otros métodos.

Ejemplo 3.12
Tomemos $f(x) = \frac{1}{x^\alpha}$, *con* $\alpha > 0$. *Entonces es fácil ver que* f *es decreciente, y por lo tanto la*

impropia $\int_1^{+\infty} \frac{1}{x^\alpha}$ *y* $\sum \frac{1}{n^\alpha}$ *son de la misma clase. En particular, esto nos permite clasificar la serie para valores de* α *en* $(1,2)$*, que era lo que nos faltaba del capítulo pasado.*

Resumiendo entonces, tanto para la integral impropia $\int_1^{+\infty} \frac{1}{x^\alpha}$ *como para la serie* $\sum \frac{1}{n^\alpha}$*, tenemos el siguiente comportamiento:*

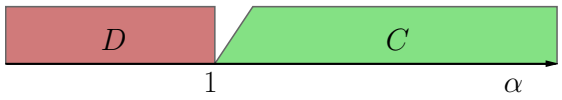


Figura 3.4: Cuando $\alpha \leqslant 1$, la impropia $\int_1^{+\infty} \frac{1}{x^\alpha} dx$ y la serie $\sum \frac{1}{n^\alpha}$ son divergentes, y cuando $\alpha > 1$ son convergentes.

Ejemplo 3.13
Clasifiquemos la serie $\sum_{n=1}^{+\infty} \frac{1}{n \log n}$*. Si tomamos la función* $f : [2,\infty) \to \mathbb{R}$*,* $f(x) = \frac{1}{x \log x}$*, tenemos que* f *es positiva y decreciente (esto último se puede ver calculando la derivada por ejemplo). Por otro lado, mediante un cambio de variable sencillo, podemos calcular la primitiva:*

$$F(x) = \int_2^x \frac{1}{x \log x} dx = \int_{\log 2}^{\log x} \frac{1}{u} du = \log u \Big|_{\log 2}^{\log x} = \log(\log(x)) - \log(\log(2)).$$

Como $\lim_{x \to +\infty} F(x) = +\infty$*, entonces la integral impropia* $\int_2^{+\infty} \frac{1}{x \log x} dx$ *es divergente, y por lo tanto también lo es la serie.*

Culminamos esta sección con el concepto de convergencia absoluta, que es completamente análogo que el visto para series.

Definición 3.14. Decimos que la integral impropia $\int_a^{+\infty} f(x)dx$ es absolutamente convergente sii $\int_a^{+\infty} |f(x)|dx$ es convergente.

Como teníamos también en el capítulo pasado, la convergencia absoluta implica la convergencia. La demostración queda como ejercicio, pues es igual que la del Teorema 2.47, *mutatis mutandis*[2].

Teorema 3.15. Si $\int_a^{+\infty} f(x)dx$ es absolutamente convergente, entonces es convergente.

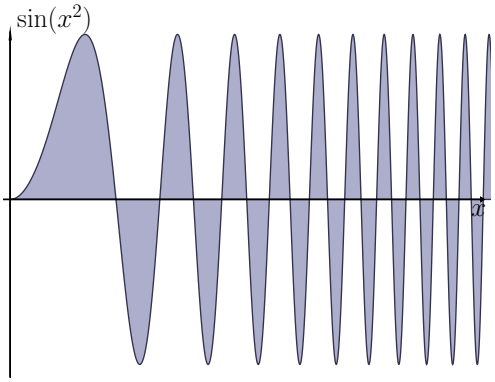
Ejercicio 3.16
Demostrar el Teorema 3.15 (ver demostración del Teorema 2.47, y adaptar lo necesario).

[2] *Mutatis mutandis* es una frase del latín que significa "cambiando lo que haya que cambiar", y en matemática se utiliza para referirse a enunciados o razonamientos que son análogos a otro. En este caso por ejemplo, se deben cambiar las reducidas enésimas por primitivas $F(x)$, el término general a_n por la función $f(x)$, etc.

Ejemplos 3.17

1. *Clasifiquemos la integral impropia* $\int_1^{+\infty} \frac{\sin x}{x^2} dx$. *Como tenemos* $\left|\frac{\sin x}{x^2}\right| \leqslant \frac{1}{x^2}$, *por comparación, la integral impropia* $\int_1^{+\infty} \frac{\sin x}{x^2} dx$ *es absolutamente convergente, y por lo tanto es convergente.*

2. *La integral* $\int_0^{+\infty} \sin(x^2) dx$ *se denomina integral de Fresnel, y tiene muchas aplicaciones en física, particularmente en óptica.*



Para clasificarla (primero entre 1 *y* $+\infty$*), escribimos* $\sin(x^2) = 2x\sin(x^2)\frac{1}{2x}$, *e integramos por partes, tomando* $2x\sin(x^2)$ *para integrar, y* $\frac{1}{2x}$ *para derivar:*

$$\int_1^{+\infty} \sin(x^2) dx = \frac{-\cos(x^2)}{2x}\Big|_1^{+\infty} - \int_1^{+\infty} \frac{\cos(x^2)}{2x^2} dx.$$

Ahora, el primer término tiene límite finito (pues el coseno está acotado, y $\frac{1}{2x}$ *tiende a cero), y la integral impropia del segundo término converge absolutamente, de manera similar al ejemplo anterior. Finalmente, la integral entre* 0 *y* 1 *no presenta problemas, ya que es la integral de una función continua en un intervalo acotado*[3]. *Por lo tanto, la integral de Fresnel es convergente.*

Ejercicio 3.18

1. *Probar que la función* $f(x) = x\cos(x^4)$ *no está acotada.*

2. *Probar que la integral impropia* $\int_0^{+\infty} x\cos(x^4)$ *es convergente.*

[3]Tuvimos que separarla dado que hicimos aparecer artificialmente un término del tipo $\frac{1}{x^2}$, que potencialmente presenta inconvenientes en cero.

3.2. Integrales impropias de segunda especie

Recordemos que para definir la integral de una función, en el curso pasado, trabajamos con funciones acotadas en intervalos acotados. Hasta ahora levantamos la segunda hipótesis, y definimos la integral en un dominio infinito. En lo que sigue, consideraremos funciones no acotadas, definidas en intervalos acotados.

Supongamos que $f:(a,b]\to\mathbb{R}$ es continua, pero no necesariamente acotada. En este caso, eso puede pasar cuando $\lim_{x\to a} f(x)$ no existe, o es infinito. Queremos definir $\int_a^b f(x)dx$, y lo haremos de la misma forma que hasta ahora: integramos en un dominio donde tenga sentido, y luego hacemos tender el dominio a $(a,b]$. Esto es:

Definición 3.19. Sea $f:(a,b]\to\mathbb{R}$ continua, y $F(x)=\int_x^b f(t)dt$. Entonces si el $\lim_{x\to a^+} F(x)=L<\infty$, decimos que la integral impropia $\int_a^b f(x)dx$ es convergente, y su valor es L. Si por el contrario el límite es infinito o no existe, decimos que la integral impropia diverge u oscila, respectivamente.

La definición cuando el dominio es $[a,b)$ es análoga.

Ejemplo 3.20
Comencemos con $\frac{1}{x^\alpha}$ en $(0,1]$. Es decir, queremos clasificar $\int_0^1 \frac{1}{x^\alpha}dx$. Calculemos entonces la primitiva:

$$F(x)=\int_x^1 \frac{1}{t^\alpha}dt=\begin{cases}\frac{1-x^{1-\alpha}}{1-\alpha} & si\ \alpha\neq 1\\ -\log(x) & si\ \alpha=1\end{cases}.$$

Por lo tanto, tenemos:

$$\lim_{x\to 0} F(x)=\begin{cases}\frac{1}{1-\alpha} & si\ \alpha<1\\ +\infty & si\ \alpha\geqslant 1\end{cases},$$

y la integral impropia $\int_0^1 \frac{1}{x^\alpha}dx$ converge solamente para $\alpha<1$.

Observar que el comportamiento es el opuesto que en infinito. Es decir, la integral impropia $\int_0^1 \frac{1}{x^\alpha}dx$ converge para $\alpha<1$, mientras que la integral $\int_1^{+\infty} \frac{1}{x^\alpha}dx$ converge para $\alpha>1$. Para $\alpha=1$, ambas son divergentes.

Tenemos los mismos resultados de comparación, equivalentes, y convergencia absoluta, enunciados para integrales impropias de primera especie.

Ejemplo 3.21
Estudiemos la integral impropia

$$\int_0^1 \frac{1}{\sqrt{x-\sin(x)}}dx$$

El único punto donde se anula el denominador es $x=0$. Para clasificar la integral, veamos el comportamiento de la función cerca del origen. Tenemos que $\sin(x) \sim x - \frac{x^3}{3}$ en cero, y por lo tanto el integrando es equivalente a $\frac{1}{x^{3/2}}$. Como el exponente de x es $\dfrac{3}{2} > 1$, la integral impropia es divergente.

Ejemplo 3.22
Sea $F(x) = \int_0^x f(t)dt$, donde f es una función continua con $f(0) = 1$. Clasifiquemos la integral impropia $\int_0^1 \frac{1}{\sqrt{F(x)}}dx$.

Nuevamente, debemos estudiar el comportamiento del integrando cerca del punto problemático, que en este caso es cero. Entonces, necesitamos el comportamiento de $F(x)$ en el origen, y para eso calcularemos su desarrollo de Taylor:

$$F(x) = F(0) + F'(0)x + o(x^2) = 0 + f(0)x + o(x^2) = x + o(x^2).$$

Tenemos entonces que $\frac{1}{\sqrt{F(x)}} \sim \frac{1}{\sqrt{x}}$ cuando $x \to 0$, y por lo tanto la integral impropia $\int_0^1 \frac{1}{\sqrt{F(x)}}dx$ es convergente.

Importante

A modo de resumen, recordemos que el comportamiento de la integral impropia $\int \frac{1}{x^\alpha}$ es opuesto en 0 y en infinito. Es decir:

$\int_0^1 \frac{1}{x^\alpha}dx$	**Converge si** $\alpha < 1$ **Diverge si** $\alpha \geqslant 1$
$\int_1^\infty \frac{1}{x^\alpha}dx$	**Converge si** $\alpha > 1$ **Diverge si** $\alpha \leqslant 1$

3.3. Integrales mixtas

Cuando en una integral aparece más de punto problemático (o dominio infinito), debemos partir la integral en suma de integrales que contengan solamente uno de esos puntos, y decimos que la integral original es convergente sii cada uno de los sumandos lo es. Así por ejemplo, si f es continua, la integral impropia $\int_{-\infty}^{+\infty} f(x)dx$ debemos escribirla como suma de $\int_{-\infty}^{a} f(x)dx$ y $\int_{a}^{+\infty} f(x)dx$, y debemos clasificar estas dos integrales. Es fácil ver que el resultado no depende de a (¡ejercicio!).

Ejemplo 3.23
La integral impropia $\int_0^{+\infty} \frac{1}{x^\alpha}dx$ tenemos que partirla en dos (pues en 0 es no acotada, y el dominio es infinito). Entonces: $\int_0^{+\infty} \frac{1}{x^\alpha}dx = \int_0^1 \frac{1}{x^\alpha}dx + \int_1^{+\infty} \frac{1}{x^\alpha}dx$.

Pero la primera converge solamente si $\alpha < 1$, y la segunda si $\alpha > 1$, por lo tanto la integral impropia $\int_0^{+\infty} \frac{1}{x^\alpha}dx$ diverge siempre (y cuando $\alpha = 1$ diverge para los dos lados).

Ejemplo 3.24
Clasificando la integral $\int_{-\infty}^{+\infty} xdx$, uno se podría ver tentado de considerar $\int_{-x}^{x} tdt$, y luego hacer tender x a infinito. Siguiendo este método, como la función identidad es impar, tendríamos que $\lim_{x\to\infty} \int_{-x}^{x} tdt = 0$. Sin embargo, hay que tener cuidado, pues siguiendo la definición, debemos partir la integral en dos sumandos: $\int_{-\infty}^{+\infty} xdx = \int_{-\infty}^{0} xdx + \int_0^{+\infty} xdx$, y estas dos integrales son divergentes, por lo tanto la impropia $\int_{-\infty}^{+\infty} xdx$ es divergente.

Ejemplo 3.25
Veamos un ejemplo que combine varios casos. Clasifiquemos la integral

$$\int_0^{+\infty} \frac{e^{-x}}{\sqrt{x}\sqrt{|2-x|}}dx.$$

Lo primero que debemos hacer es detectar dónde hay problemas de no acotación (del dominio o de la función). En este caso, tenemos una impropiedad de segunda especie en cero (por el $\frac{1}{\sqrt{x}}$), otra similar en $x = 2$ (por el $\frac{1}{\sqrt{|2-x|}}$), y finalmente en infinito. Tenemos que separar entonces en varios sumandos, donde cada uno de ellos sea una integral impropia de las estudiadas (con un solo "problema"). Esto significa que no podemos considerar la integral desde 0 hasta 2, pues tendríamos dos puntos (los extremos de integración) donde la función es no acotada. Separamos entonces:

$$\int_0^{+\infty} f(x)dx = \int_0^1 f(x)dx + \int_1^2 f(x)dx + \int_2^3 f(x)dx + \int_3^{+\infty} f(x)dx.$$

Estudiemos la primera. En el origen, el único factor problemático es $\frac{1}{\sqrt{x}}$*, pues* e^{-x} *tiende a uno, y* $\frac{1}{\sqrt{|2-x|}}$ *tiende a otro real positivo. Cerca de cero entonces, la función* $f(x)$ *es equivalente con* $\frac{1}{\sqrt{x}}$*, y por lo tanto las impropias*

$$\int_0^1 \frac{e^{-x}}{\sqrt{x}\sqrt{|2-x|}}dx \quad y \quad \int_0^1 \frac{1}{\sqrt{x}}dx$$

son de la misma clase, en este caso son convergentes.

Pasemos a la segunda. El problema aquí se encuentra en $x = 2$*. Nuevamente,* e^{-x} *tiende a una constante, y ahora* $\frac{1}{\sqrt{x}}$ *también tiende a una constante positiva. Cerca de* 2 *entonces (y por la izquierda), el integrando es equivalente a* $\frac{1}{\sqrt{2-x}}$*, y debemos clasificar* $\int_1^2 \frac{1}{\sqrt{2-x}}dx$*. Observar que la situación es equivalente a la recién clasificada:* $\frac{1}{\sqrt{2-x}}$ *cerca de* 2*, se comporta igual que* $\frac{1}{\sqrt{x}}$ *cerca de cero. De hecho, se puede hacer un cambio de variable* $(u = 2-x)$ *para pasar de una a otra. Esta integral impropia es por lo tanto convergente (ejercicio: hacer el cambio de variable y clasificar).*

El tercer sumando se clasifica igual que el segundo, y es también convergente. Pasemos al último. En infinito, la función es equivalente a $\frac{e^{-x}}{x}$*, y como la integral impropia* $\int_3^\infty \frac{e^{-x}}{x}$ *es convergente (ejercicio: clasificarla), resulta que el último sumando es también convergente.*

En resumen, todos los sumandos son integrales impropias convergentes, y por lo tanto

$$\int_0^{+\infty} \frac{e^{-x}}{\sqrt{x}\sqrt{|2-x|}}dx < \infty.$$

Capítulo 4

Topología y sucesiones en $\mathbb{R}^n$

A partir de este capítulo, vamos a intentar generalizar los conceptos de cálculo diferencial e integral en una variable (los vistos en el curso previo y algunos de los vistos en este curso) a varias variables. Con algunas definiciones y propiedades tendremos éxito, y para otras tendremos que trabajar un poco más.

Llegado este punto, conviene plantearse cuáles son los conceptos fundamentales que fueron necesarios para todas las definiciones y resultados vistos. En definitiva, eso es generalizar: abstraerse de las particularidades y quedarse con las ideas que hay de fondo.

Uno de los conceptos más importantes del cálculo diferencial es la noción de cercanía. Esto nos permitió dar una definición de límite-continuidad, derivabilidad, integrabilidad, convergencia de sucesiones, etc. Es decir, todos los conceptos fundamentales del cálculo. Básicamente cada vez que escribíamos un ε, estábamos usando la noción de cercanía, y hasta ahora, que trabajamos en $\mathbb{R}$, la distancia entre dos puntos x e y la medíamos como $|x-y|$. Veremos cómo definir formas de medir en otros conjuntos.

4.1. Topología de $\mathbb{R}^n$

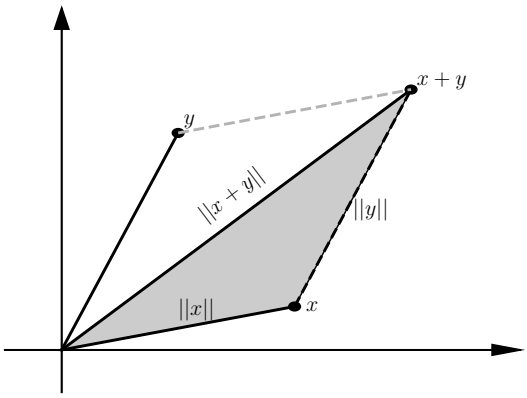
Lo primero que definiremos es la *norma* de un vector, que es una manera de medir la magnitud de un vector, la distancia al origen.

Definición 4.1. Decimos que una función $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ es una *norma* si verifica las siguientes tres propiedades:

1. $\|x\| \geqslant 0$, y $\|x\| = 0$ solamente si $x = 0$.
2. $\|\lambda x\| = |\lambda|\|x\|$, para todo $\lambda \in \mathbb{R}$.
3. $\|x+y\| \leqslant \|x\| + \|y\|$, denominada *desigualdad triangular*.

Si lo que pretendemos es medir la distancia al origen, la primera propiedad que pedimos es natural: no puede ser un valor negativo, y el único vector que tiene norma nula es el cero. Por otro lado, cuando multiplicamos el vector por un escalar λ, el vector justamente se escala, le realizamos una homotecia, entonces su norma se ve multiplicada por $|\lambda|$ (cuando λ es negativo, le cambiamos la orientación al vector).

La tercera propiedad es la más geométrica: el nombre viene del triángulo formado por el origen, el vector x, y $x+y$, donde podemos ver que la longitud del lado $x+y$ es menor que la suma de los otros dos lados, que son $\|x\|$ e $\|y\|$.



En $\mathbb{R}$, la función que usamos como norma (y que ciertamente cumple con las tres propiedades de la definición) es el valor absoluto. En los complejos, la norma usual es el módulo $|z|$.

En $\mathbb{R}^n$ la norma más usada es la euclídea, también llamada *norma dos*, y la escribimos así $\|\cdot\|_2$. Si x es un elemento de $\mathbb{R}^n$, denominaremos a sus coordenadas $(x_1, x_2, \ldots, x_n)$, y la norma euclídea se define entonces como $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2}$.

Existen varias funciones más que verifican las propiedades para ser una norma. Por ejemplo la denominada *norma uno*[1] $\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$, o en general la norma p, definida como $\|x\|_p = \sqrt[p]{|x_1|^p + |x_2|^p + \ldots + |x_n|^p}$. Otro ejemplo es la norma infinito, o norma del máximo, que se define como $\|x\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$, y puede ser pensada como el límite de la norma p cuando p tiende a infinito.

Teniendo una manera de medir normas de vectores, tenemos una forma de medir distancias entre puntos. Específicamente, podemos definir la distancia entre x e y como la norma del vector $y - x$, es decir: $d(x,y) = \|y - x\|$.

[1] También llamada norma del taxista, o norma Manhattan, debido a que, si miramos en el plano, para ir del origen a un punto x, debemos movernos solamente horizontal o verticalmente, como si fuera el cuadriculado de calles de una ciudad.

Así definida, la distancia hereda algunas propiedades de la norma. Pero en realidad lo que haremos es, al igual que hicimos arriba, definir qué tiene que cumplir una función para que la podamos considerar una función distancia[2].

Definición 4.2. Decimos que una función $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ es una *distancia* si verifica las siguientes propiedades:

1. $d(x,y) \geqslant 0$, y $d(x,y) = 0$ solamente si $x = y$.
2. $d(x,y) = d(y,x)$
3. $d(x,y) \leqslant d(x,z) + d(z,y)$, también llamada *desigualdad triangular*.

Es fácil ver que si tenemos una norma, y definimos una función distancia como $d(x,y) = \|y - x\|$, entonces esta función efectivamente verifica las tres propiedades de la definición 4.2[3].

Así, tenemos que la distancia que proviene de la norma $\|\cdot\|_2$ es la distancia euclídea, y es con la que trabajaremos en general. Sin embargo, vale la pena destacar que distintas aplicaciones requieren distintas maneras de medir distancias. La norma $\|\cdot\|_1$, la norma $\|\cdot\|_\infty$, e incluso nuevas normas definidas específicamente para problemas particulares, son muy utilizadas en aplicaciones modernas. Además, esto nos da una teoría también para medir distancias entre funciones[4] por ejemplo, que puede ser útil en otros contenidos.

La idea de definir distancias vino por la necesidad de especificar qué es "estar cerca" de un punto. En $\mathbb{R}$, usamos habitualmente el entorno de centro a y radio δ, para describir a los puntos que están a distancia menor que δ de a. Generalicemos este concepto:

Definición 4.3. Dado un punto $a \in \mathbb{R}^n$ y un número real positivo δ, llamamos *bola abierta* (o entorno, o simplemente bola) de centro a y radio δ al conjunto

$$B(a,\delta) = \{x \in \mathbb{R}^n : d(x,a) < \delta\}.$$

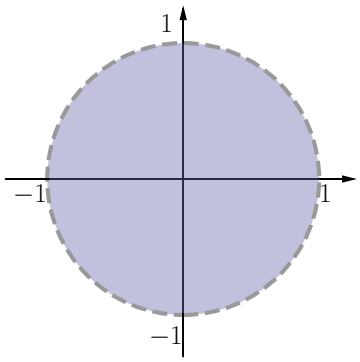
Tenemos entonces que $B(a,\delta)$ es el conjunto de puntos que se encuentra a distancia menor que δ del punto a. ¡Pero esto claramente depende de qué noción de distancia estamos usando!

[2] Cuando tenemos un conjunto $\mathscr{X}$ con una distancia $d : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$, decimos que es un *espacio métrico* $(\mathscr{X}, d)$, y parte del contenido de este curso se puede desarrollar en ese contexto más general. El lector interesado puede profundizar en [5].

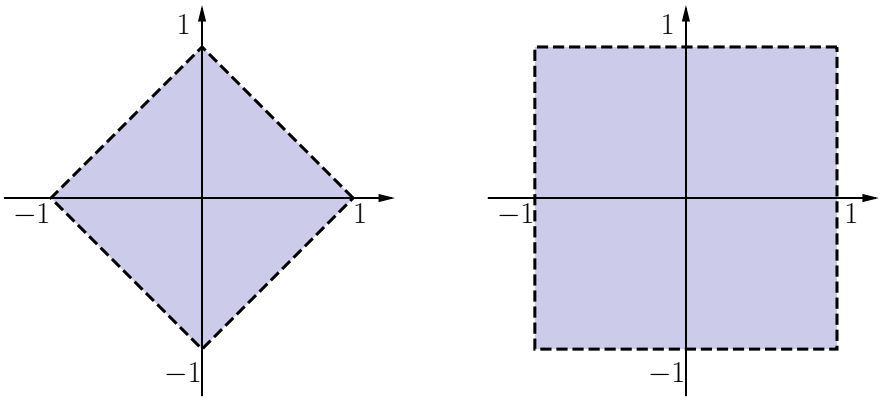
[3] Pero no todas las funciones distancia (que cumplen las tres propiedades de la definición) provienen de una norma. Por ejemplo, la distancia discreta, $d(x,y) = \begin{cases} 1 & \text{si } x \neq y \\ 0 & \text{si } x = y \end{cases}$, no proviene de ninguna norma.

[4] Es decir, el conjunto $\mathscr{X}$ es el conjunto de funciones, y se pueden definir distancias en ese conjunto. No desarollaremos esto aquí, pues estamos trabajando solamente con conjuntos de puntos en $\mathbb{R}^n$.

Cuando $n = 1$ y la distancia es el valor absoluto de la diferencia, recuperamos el entorno que conocemos. Cuando estamos en el plano (o sea, $n = 2$) y la distancia es la euclídea, la bola abierta es el interior de un círculo, sin incluir el borde. Cuando $n = 3$, es el interior de una esfera.



Ejercicio 4.4
Corroborar que las bolas de centro en el origen y radio uno, $B(0,1)$, para las distancias que provienen de las normas $\|\cdot\|_1$ y $\|\cdot\|_\infty$ son las de la figura.



De la misma forma que teníamos entornos reducidos en $\mathbb{R}$, llamamos *bola reducida* de centro a y radio δ al conjunto $B^*(a,\delta) = B(a,\delta) - \{a\}$, es decir, la bola sin el centro.

Definición 4.5. Sea A un conjunto, y denotamos por A^C su complemento. Entonces decimos que

- un punto x_0 es *interior* a A sii existe una bola $B(x_0,\delta) \subset A$ (es decir, hay una bola de centro x_0 completamente incluída en el conjunto A).
- un punto x_0 es *exterior* a A sii existe una bola $B(x_0,\delta) \subset A^C$.
- un punto x_0 es *frontera* de A sii toda bola interseca A y A^C (es decir: $\forall \delta > 0$, tenemos que $A \cap B(x_0,\delta) \neq \emptyset$, y $A^C \cap B(x_0,\delta) \neq \emptyset$).

Observar que la definición de punto frontera es equivalente a decir que x_0 no es ni interior ni exterior. Efectivamente, decir que no es interior, implica que no hay ninguna bola completamente contenida en A, es decir que todas las bolas cortan A^C. Lo mismo sucede con el hecho de no ser exterior: toda bola corta A.

Por otro lado, es claro que los puntos interiores a un conjunto necesariamente pertenecen al conjunto. De la misma forma, los puntos exteriores necesariamente pertenecen al complemento. Ahora, los puntos frontera pueden pertenecer o no al conjunto. Veamos algunos ejemplos.

Ejemplos 4.6

1. *Consideremos el conjunto $A_1 = \{(x,y) \in \mathbb{R}^2 : y \geqslant 0\}$, es decir, el semiplano superior. Entonces cualquier punto $p = (x_0, y_0)$ con $y_0 > 0$ es un punto interior. En efecto, la bola $B(p, \frac{y_0}{2})$ está completamente incluida en A. Por otro lado, todos los puntos de la forma $p = (x_0, 0)$ son puntos frontera, pues cualquier bola tendrá elementos en ambos semiplanos.*

2. *En el plano $\mathbb{R}^2$, consideremos el conjunto $A_2 = \{(x,y) \in \mathbb{R}^2 : 0 < x < 1, y = 0\}$. Entonces no hay puntos interiores, ya que no puede haber una bola de radio positivo incluida en un segmento. Todos los puntos del conjunto son puntos frontera. Además, los puntos $(0,0)$ y $(1,0)$, que no pertenecen a A, también son puntos frontera.*

3. *Consideremos finalmente el conjunto $A_3 = \{(x,y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < 1\}$. Entonces cualquier punto del conjunto es interior. En efecto, si $p = (x_0, y_0)$ es un punto de A_3, entonces si tomamos $\delta = \min\{x_0, 1 - x_0, y_0, 1 - y_0\}$, la bola de centro p y radio δ está completamente incluida en A_3 (ese δ que elegimos es la distancia más chica de p a alguno de los bordes del cuadrado).*

Cuando se da el caso de este último ejemplo, donde todos los puntos son interiores, decimos que el conjunto es abierto:

Definición 4.7. Decimos que un conjunto es abierto si todos sus puntos son interiores.

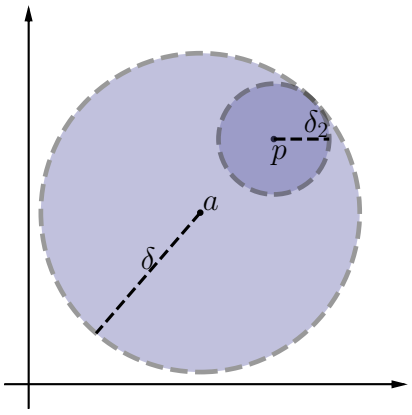
Ya tenemos un ejemplo de un conjunto abierto entonces: el conjunto A_3 del ejemplo 4.6 es abierto. El conjunto vacío también es abierto, y claramente todo el espacio $\mathbb{R}^n$ también es abierto. Veamos ahora que la que llamamos bola abierta, es efectivamente un conjunto abierto según esta última definición que vimos.

Proposición 4.8. La bola abierta $B(a, \delta)$ es un conjunto abierto.

Demostración. Para demostrar que un conjunto es abierto, debemos probar que todo punto es interior. Tomemos entonces un punto genérico $p \in B(a, \delta)$, y encontremos una bola de centro p, completamente incluida en $B(a, \delta)$.

La intuición geométrica nos dice que basta considerar la distancia de p al borde de la bola. Esto es, consideremos $\delta_2 = \delta - d(a,p)$. Veamos que con esta elección, tenemos efectivamente que $B(p,\delta_2) \subset B(a,\delta)$. Para eso, tomemos un punto $x \in B(p,\delta_2)$, y veamos que también está en $B(a,\delta)$, es decir, que $d(a,x) < \delta$. Finalmente, por la desigualdad triangular tenemos

$$d(a,x) \leqslant d(a,p) + d(x,p) < d(a,p) + \delta_2 = \delta.$$



□

El concepto importante aquí es que no importa cuan cerca del borde esté el punto p. Si está en $B(a,\delta)$ (es decir, la distancia de p a a es *estrictamente* menor que δ), entonces hay espacio para encontrar una bola de centro p, que quede completamente incluida en $B(a,\delta)$.

Observemos que la definición de punto interior (y por lo tanto de conjunto abierto), se basa en la definición de bola abierta, que depende fuertemente del espacio ambiente en el que estamos trabajando ($\mathbb{R}^n$). Es decir, si miramos el intervalo $(0,1)$ como subconjunto de $\mathbb{R}$, es un conjunto abierto. Ahora, si miramos el conjunto A_2 del ejemplo 4.6, que es una copia del intervalo $(0,1)$, pero ahora como subconjunto de $\mathbb{R}^2$, entonces el conjunto no es abierto.

Volvamos a mirar ejemplos de conjuntos abiertos en $\mathbb{R}$. Claramente todos los intervalos abiertos (a,b) son efectivamente conjuntos abiertos, pero, ¿son los únicos?. Una manera fácil de construir otros ejemplos (que no sean intervalos), es uniendo conjuntos abiertos. Así, por ejemplo, el conjunto $(0,1)\cup(2,3)$ es abierto. ¿Qué sucede si unimos una cantidad infinita de conjuntos abiertos? ¿Y con la intersección? Veamos estas propiedades.

Proposición 4.9. Si A y B son dos conjuntos abiertos, entonces $A \cap B$ es abierto.

Demostración. Tomemos un punto p en la intersección, y veamos si es interior a $A \cap B$.

Como $p \in A$, entonces es interior a A (pues A es abierto). Por lo tanto existe un $\delta_A > 0$ tal que $B(p,\delta_A) \subset A$. De la misma forma, como $p \in B$, existe un $\delta_B > 0$ tal que $B(p,\delta_B) \subset B$.

Tenemos entonces dos bolas centradas en p, una incluida en A, y otra incluida en B. Pero si tomamos el más chico de los dos radios, entonces la bola resultante estará en ambos conjuntos. Esto es, tomemos $\delta = \min\{\delta_A, \delta_B\}$, y entonces tenemos que $B(p,\delta) \subset A$ y $B(p,\delta) \subset B$, es decir, $B(p,\delta) \subset A \cap B$, como queríamos probar. □

Observemos primero que esta propiedad se extiende trivialmente para cualquier intersección finita de conjuntos abiertos[5], pero al menos esta demostración no sirve para una cantidad infinita de conjuntos: en un momento tomamos el mínimo de los radios δ_A y δ_B, pero si tuviéramos una cantidad infinita de radios (uno por cada conjunto), entonces el mínimo no tiene por qué existir. Dar un ejemplo entonces de una cantidad infinita de conjuntos abiertos, tal que su intersección no sea abierta.

Con la unión, sin embargo, tenemos mejor suerte. En el próximo resultado, consideraremos la unión de una cantidad arbitraria de conjuntos, que indexaremos según un conjunto de índices L. Si tenemos solamente dos conjuntos, entonces $L = \{1,2\}$, pero podemos unir una cantidad infinita de conjuntos, y en ese caso tendremos por ejemplo $L = \mathbb{N}$, o incluso otro conjunto infinito "más grande".

Proposición 4.10. Sea $(A_\lambda)_{\lambda\in\mathscr{L}}$ una familia de conjuntos abiertos, $A_\lambda \subset \mathbb{R}^n$. Entonces su unión $\bigcup_\lambda A_\lambda$ es un conjunto abierto.

Demostración. Para demostrar que la unión es abierta, debemos considerar un punto genérico en la unión, y demostrar que es interior. Sea entonces $p \in \bigcup_\lambda A_\lambda$. Como p está en la unión, entonces pertenece a alguno de los A_λ, digamos que $p \in A_{\lambda_0}$. Pero como A_{λ_0} es abierto, entonces existe δ tal que $B(p,\delta) \subset A_{\lambda_0}$. Ahora, como A_{λ_0} es uno de los conjuntos que participan en la unión, entonces $B(p,\delta) \subset A_{\lambda_0} \subset \bigcup_\lambda A_\lambda$, como queríamos probar. □

Recordemos que los puntos frontera de un conjunto podían pertenecer o no al conjunto. Pero si tenemos un conjunto abierto, entonces los puntos frontera *no pueden* pertenecer al conjunto (ya que todos los puntos del conjunto son interiores). En el otro extremo, tenemos los conjuntos donde *todos* los puntos frontera pertenecen al conjunto, que son los conjuntos cerrados:

Definición 4.11. Decimos que un conjunto es cerrado si su complemento es abierto.

Resta ver que esta definición es coherente con la intuición que dimos sobre la frontera del conjunto.

Ejercicio 4.12

1. *Demostrar que si x_0 es un punto frontera de un conjunto A, entonces también es frontera de A^C. Es decir, un conjunto y su complemento, comparten la frontera.*

2. *Probar que si C es cerrado y x_0 es un punto frontera de C, entonces el punto x_0 pertenece a C (Sugerencia: utilizar la primera parte de este ejercicio).*

[5]Una manera fácil de hacer esto es ir intersecando los conjuntos de a uno. Pero también se puede adaptar la prueba: hay que considerar n bolas, una para cada conjunto, y luego tomar el mínimo de los n radios.

3. *Recíprocamente, probar que si un conjunto C contiene a todos sus puntos frontera, entonces es cerrado.*

Ejemplos 4.13

1. *El conjunto formado por un punto $A = \{p\}$ es cerrado.*
2. *El conjunto $\overline{B(p,\delta)} = \{x \in \mathbb{R}^n : d(p,x) \leqslant \delta\}$ es un conjunto cerrado.*
3. *La esfera $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ es un conjunto cerrado. Cuando $n = 2$, obtenemos la circunferencia unidad.*
4. *El conjunto A_1 del ejemplo 4.6 es un conjunto cerrado.*

Ejercicio 4.14
Corroborar que los conjuntos del ejemplo anterior son cerrados.

Es importante observar que hay conjuntos que no son abiertos ni cerrados, como por ejemplo un intervalo $[0,1)$ en $\mathbb{R}$. También hay conjuntos que son abiertos **y** cerrados. En $\mathbb{R}^n$ solamente hay dos: el vacío y todo $\mathbb{R}^n$.

En el capítulo 2 ya trabajamos sobre algunas cuestiones topológicas, en particular con puntos de acumulación. Vamos ahora a definir nuevamente (aunque es una copia de la definición 2.22) punto de acumulación, en este nuevo contexto:

Definición 4.15. Dado un conjunto $A \subset \mathbb{R}^n$, decimos que $p \in \mathbb{R}^n$ es un punto de acumulación de A sii toda bola reducida corta al conjunto. Es decir, si $\forall \varepsilon > 0$, $B^*(p,\varepsilon) \cap A \neq \emptyset$

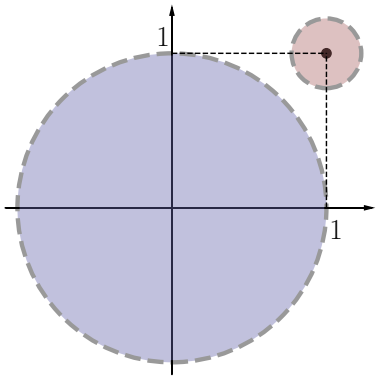
A primera vista, la definición parece tener ciertas similitudes con la de puntos frontera de un conjunto. Vale la pena estudiar sus diferencias y parecidos.

El concepto de punto de acumulación es, como ya vimos en el capítulo 2, que nos podemos acercar al punto tanto como queramos con elementos del conjunto. Esto es cierto, para empezar, para todos los puntos interiores. Si vamos a la definición de punto de acumulación, efectivamente si tenemos punto interior, cualquier bola reducida tiene intersección no vacía con el conjunto (pues hay una bola *completamente* contenida en el conjunto, por lo tanto, por más que le quitemos el centro, seguirá teniendo infinitos puntos del conjunto).

Ejemplo 4.16
Consideremos el conjunto $B(0,1)$, del que ya vimos que todos sus puntos son interiores. Por lo tanto, todos sus puntos son de acumulación, y esto tiene sentido: cualquier punto del conjunto está "rodeado" por el conjunto $B(0,1)$. En este caso, además, la frontera (que es la circunferencia unidad), también está compuesta de puntos de acumulación. Efectivamente, nos podemos acercar a cualquier punto de la circunferencia con puntos de la bola. Dicho con la terminología de la definición, cualquier bola reducida centrada en un punto frontera, tiene intersección no vacía con $B(0,1)$.

Volvamos al caso general. Para un punto frontera, cualquier bola tiene intersección no vacía con el conjunto (también con el complemento, pero aquí nos interesa la intersección con el conjunto), y por lo tanto hay al menos un punto del conjunto en esa intersección. El problema surge cuando tenemos que mirar la bola reducida, pues quizás quitamos al único punto que aparecía en la intersección. Veamos este comportamiento en el siguiente ejemplo.

Ejemplo 4.17
Consideremos el conjunto $A = B(0,1) \cup \{(1,1)\}$, que surge de unir el punto $(1,1)$ al conjunto del ejemplo anterior. El punto $(1,1)$ es un punto frontera, pero "no tiene del conjunto a nadie alrededor". Efectivamente, si tomamos una bola reducida de centro $(1,1)$ y radio chico, como en la figura, no contiene ningún punto del conjunto A.



En general, estos puntos p, donde existe una bola $B(p,\delta)$ tal que el único punto del conjunto en esa bola, es p, se llaman *puntos aislados*, y son puntos frontera, pero no de acumulación.

Veamos ahora una caracterización de los conjuntos cerrados, en función de sus puntos de acumulación, sobre todo para ejercitar los conceptos y la forma de demostrar estas propiedades.

Proposición 4.18. Un conjunto C es cerrado sii C contiene a todos sus puntos de acumulación.

Demostración. ($\Rightarrow$) Sea p un punto de acumulación de C. Queremos probar que $p \in C$.

Supongamos que no, es decir, que $p \in C^C$. Pero como C es cerrado por hipótesis, entonces C^C es abierto, y por lo tanto p es interior a C^C. Es decir, existe un $\delta > 0$ tal que $B(p,\delta) \subset C^C$. Pero esto no puede suceder, ya que p es de acumulación de C (y por lo tanto todo entorno reducido debe intersecar al conjunto). El absurdo provino de suponer que $p \notin C$, por lo que queda demostrado el resultado en la primera dirección.

($\Leftarrow$) Para demostrar que C es cerrado, debemos probar que su complemento es abierto. Sea entonces $p \in C^C$, y queremos ver que es interior a C^C. Ahora bien, si p no está en C, entonces no puede ser un punto de acumulación de C (ya que por hipótesis, C contiene a todos sus puntos de acumulación). Es decir que es falso que toda bola reducida corta a C, o, dicho de otra manera, existe una bola $B^*(p,\delta)$ que no corta a C. Pero si p no está en C, y $B^*(p,\delta) \cap C = \emptyset$, entonces la bola $B(p,\delta)$ está completamente incluida en C^C, y por lo tanto p es interior a C^C. Así, C^C es abierto, y entonces C es cerrado, como queríamos probar. □

Dado un conjunto A, los conjuntos de puntos interiores, frontera, de acumulación, tienen nombres y propiedades particulares. Por ejemplo, el conjunto formado por todos sus puntos interiores es un conjunto abierto, y de hecho es el abierto más grande contenido en A. No

estudiaremos estos resultados aquí, pero sí veremos la denominación de estos conjuntos, para trabajar con más comodidad.

Notación:

- Llamamos $\mathring{A}$ al conjunto de puntos interiores de A.
- Llamamos ∂A al conjunto de puntos frontera de A.
- Llamamos A' al conjunto de puntos de acumulación de A.
- Definimos como $\overline{A} = A \cup \partial A$ al conjunto clausura de A.

Con esta notación, se puede re-escribir la definición de conjunto abierto, de la siguiente manera: un conjunto A es abierto sii $A = \mathring{A}$. También es cierto que C es cerrado sii $\overline{C} = C$.

Ejemplo 4.19
En el plano, consideremos el siguiente conjunto

$$A_1 = \left\{ x \in \mathbb{R}^2 : y = 0, x = \frac{1}{n}, n \in \mathbb{N} \right\}.$$

Entonces, A_1 no tiene puntos interiores, y por lo tanto no es abierto. Todos los puntos de A_1 son frontera, pero, ¿hay otros?.

El $(0,0)$ no pertenece al conjunto, pero todo entorno $B(0,\delta)$ contiene puntos del conjunto (y claramente también puntos del complemento). El conjunto de puntos frontera es entonces $\partial A_1 = A_1 \cup \{(0,0)\}$.

Con el mismo argumento, vemos que $(0,0)$ es un punto de acumulación de A_1, y además es el único, ya que todos los puntos del conjunto son puntos aislados (¿por qué?). El conjunto A_1 entonces no es cerrado, ya que no contiene a todos sus puntos de acumulación.

Si consideramos ahora el conjunto $A_1 \cup \{(0,0)\}$, entonces este conjunto sí es cerrado (y en realidad no es otro que $\overline{A_1}$).

Algunos de los resultados más importantes de funciones continuas, se basan mucho en la geometría del dominio. Tenemos por ejemplo, que toda función continua en un intervalo $[a,b]$ es uniformemente continua, o que tiene máximo y mínimo (Weierstrass). Es fácil ver que estos resultados no son ciertos si el dominio en todo $\mathbb{R}$, o un intervalo (a,b), por ejemplo. Entonces, ¿qué es lo que realmente importa? ¿cómo podemos generalizar estos resultados cuando trabajemos con funciones en $\mathbb{R}^n$?. Es decir, lo que estamos buscando es el análogo al intervalo $[a,b]$ pero en $\mathbb{R}^n$, para poder decir algo sobre le comportamiento de las funciones definidas en esos dominios.

Pasemos entonces a las últimas definiciones de esta naturaleza.

Definición 4.20. Decimos que un conjunto A es acotado si existe $K > 0$ tal que $A \subset B(0,K)$. Es decir, si podemos incluir al conjunto en alguna bola.

Observemos primero que, en $\mathbb{R}$, esta definición coincide con la que ya teníamos de conjuntos acotados. En particular los intervalos $[a,b]$ son acotados. Por otro lado, la diferencia con un intervalo de la forma (a,b), donde fallan los teoremas mencionados para funciones continuas, es que estos últimos no contienen a sus puntos frontera.

Definición 4.21. Decimos que un conjunto es compacto si es cerrado y acotado[6].

Los intervalos $[a,b]$ son conjuntos compactos en $\mathbb{R}$. ¡Pero no son los únicos! Piense un ejemplo de un conjunto compacto que no sea un intervalo.

Ejemplos 4.22

1. *La bola cerrada $\overline{B(0,1)}$ es un conjunto compacto.*
2. *El conjunto $A_1 \cup \{(0,0)\}$ del ejemplo 4.19, es compacto.*
3. *Más en general, si A es un conjunto acotado, entonces $\overline{A}$ es compacto.*

4.2. Sucesiones en R^n

Cuando trabajamos con sucesiones en el capítulo 2, dijimos que lo que realmente importaba en la definición era que el dominio de la función fueran los naturales. La extensión de la definición a sucesiones en $\mathbb{R}^n$ es inmediata.

Definición 4.23. Una sucesión es una función $a : \mathbb{N} \to \mathbb{R}^n$.

Es decir, ahora es una lista de puntos en $\mathbb{R}^n$.

Como antes, denotaremos con a_k tanto a la sucesión como al elemento número k. Ahora, además, tenemos otro elemento para tener cuidado, ya que cada elemento de la sucesión es un punto en $\mathbb{R}^n$, y por lo tanto tiene n coordenadas, es decir, $a_k = (a_{k,1}, a_{k,1}, \ldots, a_{k,n})$.

Con todas las generalizaciones que hicimos en lo que va del capítulo, la definición de convergencia de una sucesión es prácticamente un calco de la definición 2.3 que ya dimos.

Definición 4.24. Decimos que la sucesión a_n tiene límite $L \in \mathbb{R}^n$, y lo denotamos $\lim_{n\to\infty} a_n = L$ sii $\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}$ tal que $\forall n \geqslant n_0$ se tiene que $a_n \in B(L,\varepsilon)$.

[6]Existe una definición más general de compacidad, que es válida en otros espacios. En $\mathbb{R}^n$ esta definición coincide con la que dimos, por lo que nos limitaremos a trabajar con esta definición.

Observemos mediante el siguiente ejercicio que esta definición es equivalente a pedir que la sucesión (de números reales) $d(a_n, L)$ tienda a cero con n.

Ejercicio 4.25
Escriba qué quiere decir que esta sucesión de números reales $d(a_n, L)$ tienda a cero (utilizando la definición para sucesiones reales, que vimos en el capítulo 2), y compárelo con la definición 4.24.

Proposición 4.26. Sea a_k una sucesión en $\mathbb{R}^n$ y $L = (L_1, L_2, \ldots, L_n)$. Entonces,

$$\lim_{k\to\infty} a_k = L \iff \lim_{k\to\infty} a_{k,i} = L_i \text{ para todo } i = 1, 2, \ldots n.$$

Es decir, una sucesión en $\mathbb{R}^n$ converge a L si y sólo si cada una de sus coordenadas converge a la coordenada correspondiente de L.

Demostración. Hagamos la demostración para la norma $\|\cdot\|_2$.

($\Rightarrow$) Observemos que $0 \leqslant |a_{k,i} - L_i| \leqslant \|a_k - L\|_2$. Ahora, la sucesión $\|a_k - L\|_2$ tiende a cero, pues $\lim a_k = L$. Por lo tanto, para cada coordenada, tenemos que la sucesión $|a_{k,i} - L_i|$ tiende a cero.

($\Leftarrow$) Ahora tenemos que $|a_{k,i} - L_i|$ tiende a cero para cada i. Entonces, si miramos la expresión de $\|a_k - L\|_2$:

$$\|a_k - L\|_2 = \sqrt{(a_{k,1} - L_1)^2 + (a_{k,2} - L_2)^2 + \ldots + (a_{k,n} - L_n)^2},$$

vemos que cada término bajo la raíz, lo podemos hacer tan chico como queramos, y por lo tanto también $\|a_k - L\|_2$. Es un ejercicio formalizar esto. Es decir: tomar un ε arbitrario, hacer cada uno de los términos $|a_{k,1} - L_1|$ más chicos que una cantidad conveniente (a partir de cierto k_{0_i} diferente para cada coordenada eventualmente), y luego elegir un k_0 a partir del cual se tenga $\|a_k - L\|_2 < \varepsilon$. □

Ejemplos 4.27

1. *La sucesión de $\mathbb{R}^2$ definida como $a_k = \left(\frac{1}{k}, k^2\right)$ no es convergente, ya que su segunda componente no lo es.*
2. *La sucesión de $\mathbb{R}^3$ definida como $a_k = \left(\sqrt{k}, 1, \frac{1}{k}\right)$ no es convergente, ya que su primera componente no lo es.*
3. *La sucesión de $\mathbb{R}^3$ definida como $a_k = \left(2, \frac{(-1)^k}{k}, 1 + e^{-k}\right)$ converge al punto $(2, 0, 1)$.*

De la última proposición se desprenden fácilmente varias propiedades, como por ejemplo las listadas aquí:

Proposición 4.28. Sean a_k y b_k sucesiones de $\mathbb{R}^n$ tales que $\lim a_k = a$ y $\lim b_k = b$, y λ_k una sucesión de números reales convergente a $\lambda \in \mathbb{R}$. Entonces

1. $\lim a_k + b_k = a + b$
2. $\lim \lambda_k a_k = \lambda a$
3. $\lim \|a_k\| = \|a\|$

El comportamiento de las sucesiones, en el sentido de qué puede pasar con sus límites, dice mucho sobre el conjunto o el espacio en donde se encuentra. Por ejemplo, en $\mathbb{R}$ las sucesiones pueden tener límite infinito, pero si tenemos una sucesión en el $[0,1]$, debe tener una subsucesión convergente. Relacionemos entonces algunos conceptos sobre conjuntos con el comportamiento de sucesiones.

Si tenemos una sucesión a_k incluida en un conjunto A (es decir que $a_k \in A$ para todo k), y que además es convergente a un punto a, entonces, ¿qué podemos decir sobre este punto?. Consideremos la sucesión $a_k = \frac{1}{k}$, que está contenida en el conjunto $A = (0,2)$. Su límite, cero, no pertenece al conjunto. Es decir, tenemos una sucesión completamente contenida en A, pero que tiende a un punto que está fuera del conjunto. ¿En qué conjuntos puede pasar esto?

Decimos que un conjunto A es *cerrado por sucesiones* cuando tiene la siguiente propiedad: si una sucesión de elementos del conjunto $\{a_k\} \subset A$ es convergente a $\lim a_k = a$, entonces $a \in A$. Es decir, es un conjunto que contiene todos los límites de las sucesiones de elementos del conjunto.

Veamos que esta propiedad es equivalente a ser cerrado.

Proposición 4.29. Un conjunto C es cerrado sii es cerrado por sucesiones.

Demostración. ($\Rightarrow$) Tenemos primero que C es cerrado. Sea entonces una sucesión cualquiera a_k de elementos de C, tal que $\lim a_k = a$. Queremos probar que $a \in C$.

Supongamos que no. Es decir, que $a \in C^C$. Pero si C es cerrado, C^C es abierto, y por lo tanto existe una bola $B(a,\varepsilon)$ completamente contenida en C^C. Pero como $\lim a_k = a$, para ese valor de ε, existe un k_0 a partir del cual la sucesión se encuentra en $B(a,\varepsilon)$. Ahora, la sucesión está compuesta por elementos del conjunto, y teníamos que $B(a,\varepsilon) \subset C^C$, por lo que llegamos a un absurdo.

($\Leftarrow$) Debemos probar ahora que C es cerrado, sabiendo que cumple esa propiedad sobre las sucesiones. Lo que haremos, utilizando lo visto en el ejercicio 4.12, es tomar un punto frontera de C, y demostrar que efectivamente pertenece al conjunto.

Sea entonces x_0 un punto frontera de C. La idea será construir una sucesión de elementos del conjunto C, que sea convergente a x_0. Si x_0 es frontera, entonces en toda bola $B(x_0,\varepsilon)$ hay puntos de C. Podemos entonces ir eligiendo valores de ε cada vez más chicos, y tomando

un punto de C en cada una de esas bolas. Formalmente, tomemos $\varepsilon_k = \frac{1}{k}$, y como en la bola $B(x_0, \varepsilon_k)$ hay por lo menos un punto de C, tomamos ese punto y le llamamos a_k. De esta manera, construimos una sucesión a_k de elementos de C, que converge a x_0 por construcción.

Ahora, como C es cerrado por sucesiones, y $a_k \in C$ para todo k, tenemos que $x_0 \in C$, como queríamos probar. □

La última noción que vimos en la sección anterior fue la de compacidad. Para relacionar este concepto con sucesiones, necesitamos trabajar con subsucesiones, como ya lo hicimos en el capítulo 2. De hecho la definición de subsucesión es una copia de la que ya hicimos, dado que lo importante no es el espacio de llegada ($\mathbb{R}$ antes, $\mathbb{R}^n$ ahora).

Definición 4.30. Sea a_k una sucesión en $\mathbb{R}^n$, y k_j una sucesión estrictamente creciente de números naturales. Entonces la sucesión a_{k_j} es una subsucesión de a_k.

Es sencillo corroborar que si a_k es una sucesión convergente, entonces toda subsucesión de a_k también lo es, y además convergen al mismo límite. También es cierto que toda sucesión acotada tiene una subsucesión convergente, pero la demostración no se desprende trivialmente del resultado que teníamos en los reales.

Supongamos que tenemos una sucesión acotada en $\mathbb{R}^2$, que llamaremos (x_k, y_k). Entonces, cada una de las sucesiones reales, x_k e y_k, es acotada. Nos vemos tentados a aplicar el teorema para $\mathbb{R}$, en ambas coordenadas, para luego combinar las dos subsucesiones extraídas, formando una subsucesión de (x_k, y_k). El problema es que estas subsucesiones fueron extraídas independientemente, y por lo tanto pueden no tener índices en común. Por ejemplo, quizás la subsucesión convergente de x_k es la de los índices impares, y la de y_k la de los múltiplos de 4. Entonces, ¿cuál sería una subsucesión convergente de (x_k, y_k)?. Debemos obtener las subsucesiones con un poco más de cuidado.

Teorema 4.31. Toda sucesión acotada de $\mathbb{R}^n$ tiene una subsucesión convergente.

Demostración. Sea $a_k = (a_{k,1}, a_{k,2}, \ldots, a_{k,n})$ una sucesión acotada de $\mathbb{R}^n$, y consideremos primero la sucesión real de la primera coordenada, $a_{k,1}$.

Como a_k es acotada, también lo es $a_{k,1}$. Pero esta sucesión es real (es la primera coordenada de a_k), y por lo tanto tiene una subsucesión convergente, pues para sucesiones en $\mathbb{R}$ ya demostramos el resultado en 2.26.

Para comodidad con la notación, sea $N_1 \subset \mathbb{N}$ el conjunto de índices que define esta subsucesión[7]. Escribiremos entonces $(a_{k,1})_{k \in N_1}$ para referirnos, en este caso, a la subsucesión convergente de $a_{k,1}$ extraída gracias al teorema 2.26.

[7] Es decir, los infinitos índices que elegimos en la subsucesión. Por ejemplo, si la subsucesión fuera la de los índices pares, a_{2k}, entonces N_1 sería el conjunto de los números pares.

Pasemos ahora a la segunda coordenada. Consideremos, en lugar de la sucesión entera $a_{k,2}$, la subsucesión que se obtiene al considerar los índices N_1 elegidos en la primera coordenada, es decir $(a_{k,2})_{k\in N_1}$. Esta subsucesión es acotada, pero no necesariamente convergente. Lo que sí podemos afirmar es que contiene una subsucesión convergente (nuevamente, aplicando el resultado para sucesiones reales). Llamemos $N_2 \subset N_1 \subset \mathbb{N}$ al conjunto de índices que determina esta subsucesión.

Tenemos ahora, entonces, la subsucesión convergente $(a_{k,2})_{k\in N_2}$. Pero además, la subsucesión de la primera coordenada $(a_{k,1})_{k\in N_2}$, considerando este último conjunto de índices, también es convergente, pues es una subsucesión de $(a_{k,1})_{k\in N_1}$, que era convergente.

Repetimos el argumento para las n coordenadas, y obtenemos un conjunto infinito de índices N_n, tal que $(a_{k,n})_{k\in N_n}$ es convergente. Pero como este conjunto de índices N_n lo fuimos eligiendo cuidadosamente, anidando los índices de las subsucesiones convergentes para cada coordenada, cada una de las subsucesiones $(a_{k,i})_{k\in N_n}$ es convergente, y por lo tanto encontramos una subsucesión de a_k convergente, que es $(a_k)_{k\in N_n}$. □

El siguiente corolario, que se desprende muy fácilmente del resultado anterior, será clave para demostrar algunas propiedades de las funciones continuas en el siguiente capítulo.

Corolario 4.32. Si $K \subset \mathbb{R}^n$ es un conjunto compacto, y a_k es una sucesión de elementos de K, entonces a_k posee una subsucesión convergente, y además su límite es un elemento de K.

Demostración. Como K es compacto, en particular es acotado. Por lo tanto la sucesión a_k es acotada, y en virtud del Teorema 4.31 recién demostrado, contiene una subsucesión convergente. Llamemos $(a_k)_{k\in N_1}$ a esa subsucesión, y sea a su límite.

Ahora, esta subsucesión convergente también está en K, que es cerrado, por lo tanto, como vimos en la Proposición 4.29, su límite a pertenece a K, lo que concluye la prueba. □

En realidad, esta propiedad de un conjunto (que toda sucesión de elementos del conjutno tenga una subsucesión convergente a un punto del conjunto) es equivalente a pedir que le conjunto sea compacto. Es una *caracterización* de los conjuntos compactos.

Si bien no demostraremos esta equivalencia, las ideas que hay atrás son las que intenta explotar el siguiente ejercicio.

Ejercicio 4.33

1. *Elija un conjunto C cerrado pero no acotado. Encuentre una sucesión a_k de elementos de C que no contenga ninguna subsucesión convergente a un elemento de C.*

2. *Elija un conjunto A acotado pero no cerrado. Encuentre una sucesión a_k de elementos de A que no contenga ninguna subsucesión convergente a un elemento de A.*

Capítulo 5

Continuidad

En el capítulo anterior comenzamos con algunas generalizaciones. Definimos sucesiones en $\mathbb{R}^n$, y tenemos una manera de medir cercanías en $\mathbb{R}^n$, tenemos una topología. En este capítulo comenzaremos a generalizar las funciones, que hasta el momento eran funciones de una variable, tomando valores reales.

Comencemos pensando cuáles son los aspectos importantes que conocemos de funciones de una variable, para ver qué aspiramos generalizar en lo que resta del curso. Todas las propiedades y resultados más importantes que estudiamos de funciones en una variable se engloban dentro de tres grandes áreas: continuidad, derivabilidad e integrabilidad. Veremos qué elementos se pueden generalizar.

Algunas de estas generalizaciones serán fáciles, y otras un poco más complicadas. La generalización de algunos resultados serán ciertas, y otras no. Comencemos tomando intuición geométrica y operativa de funciones de varias variables.

5.1. Funciones en $\mathbb{R}^n$

Recordemos que una función consta de tres elementos: un dominio, un codominio, y una regla que a cada elemento del dominio le asigna uno del codominio. Trabajaremos en este capítulo fundamentalmente con funciones con dominio en $\mathbb{R}^n$ (usando muchas veces $\mathbb{R}^2$ para ejemplificar) y codominio $\mathbb{R}$[1]. Es decir, funciones $f : \mathbb{R}^n \to \mathbb{R}$.

Ya conocemos algunos ejemplos de estas funciones:

[1] Luego sí veremos funciones con codominio $\mathbb{R}^m$.

Ejemplos 5.1

1. *Las normas con las que trabajamos el capítulo anterior son funciones que toman valores reales (positivos específicamente).*

$$f_1(x,y) = \|(x,y)\|_1 = |x| + |y|.$$

$$f_2(x,y) = \|(x,y)\|_2 = \sqrt{x^2+y^2}.$$

2. *Otras clase de funciones conocidas son las transformaciones lineales. En particular, un ejemplo de transformación lineal de $\mathbb{R}^2$ a $\mathbb{R}$ podría ser*

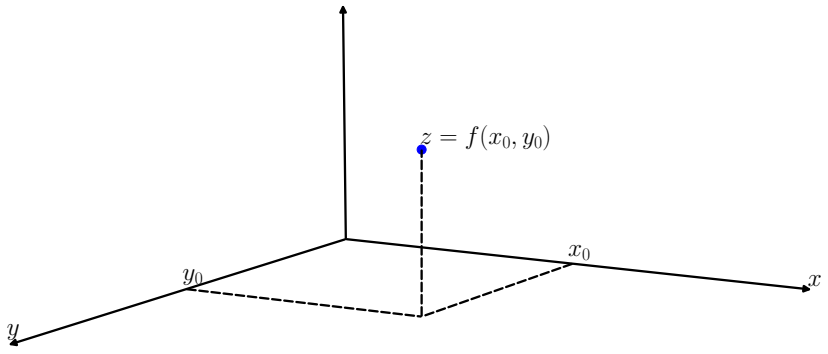
$$f_3(x,y) = 2x - 3y.$$

3. *Otro ejemplo de función definida en el plano, de forma arbitraria, es*

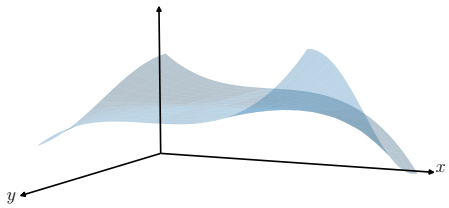
$$f_4(x,y) = e^{x^2 y} + \sin(x+y) + 1.$$

Veamos cómo representar gráficamente estas funciones. El gráfico de una función es un subjconjunto del producto cartesiano del dominio y codominio. Es decir, cuando trabajamos con funciones de una variable, $f : \mathbb{R} \to \mathbb{R}$, representamos su gráfico en el plano $\mathbb{R} \times \mathbb{R}$. El gráfico de $f(x) = x^2$ son los puntos del plano cuyas coordenadas (x,y) cumplen que la segunda coordenada es igual a la primera al cuadrado.

Ahora, si queremos "dibujar" una función $f : \mathbb{R}^2 \to \mathbb{R}$, lo debemos hacer en el espacio tridimensional $\mathbb{R}^2 \times \mathbb{R}$. Si las coordenadas de este $\mathbb{R}^3$ son (x,y,z), usaremos el "piso", es decir el plano (x,y), para representar el dominio, y la dimensión restante, la altura z, para representar los valores funcionales $f(x,y)$.



Con funciones de $\mathbb{R}$ a $\mathbb{R}$, el gráfico de la función, que se puede pensar como "unir" los puntos del plano de la forma $(x, f(x))$, resulta una curva, una figura de "dimensión uno".Cuando trabajamos con funciones de $\mathbb{R}^2$ a $\mathbb{R}$, este gráfico resulta una superficie, una figura de "dimensión dos".



Observemos que si quisiéramos representar una función $f : \mathbb{R}^3 \to \mathbb{R}$, tendríamos que graficarla en $\mathbb{R}^4$, por lo que no vamos a poder representar funciones reales cuyo dominio sea de dimensión mayor que dos.

Veamos ahora una herramienta para obtener información parcial sobre la forma del gráfico de una función.

Conjuntos de nivel

Los conjuntos de nivel son subconjuntos del dominio, cuyos puntos tienen la misma imagen por la función.

Dado un número $k \in \mathbb{R}$, el conjunto de nivel k es $\quad C_k = \{(x,y) \in \mathbb{R}^2 : f(x,y) = k\}$.

Es decir, son todos los puntos del dominio, cuya imagen es exactamente k. Dicho de otra manera, es el conjunto de preimágenes de k. Gráficamente, corresponde a ver los puntos en la intersección del gráfico de la función, con el plano horizontal de altura k (o sea, $z = k$).

Tomemos por ejemplo la función $f(x,y) = x^2 + y^2$. La curva de nivel $k = 1$ son los puntos (x,y) del plano, tales que $f(x,y) = 1$, es decir, tales que $x^2 + y^2 = 1$. Toda la circunferencia unidad entonces tiene como imagen el uno, y por lo tanto si intersecamos el gráfico de la función con un plano horizontal, paralelo al plano (x,y), a la altura $z = 1$, la intersección será una circunferencia de radio uno.

Todas las curvas de nivel de esta función son circunferencias concéntricas.

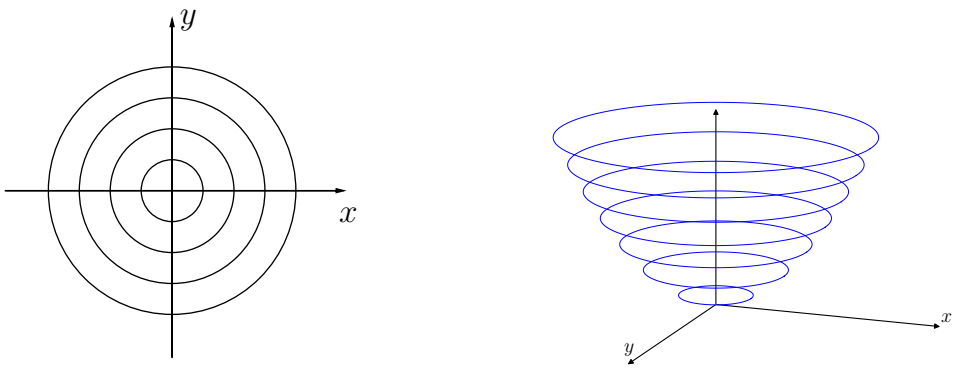


Figura 5.1: Curvas de nivel de $f(x,y) = x^2 + y^2$. A la izquierda, el dominio de la función, y algunas curvas de nivel. A la derecha, las curvas de nivel a la altura correspondiente.

Si bien el conjunto de todas las curvas de nivel contienen toda la información para determinar la función, no hay que apresurarse a sacar conclusiones a partir de una vista rápida de algunas curvas de nivel. Por ejemplo, si tomamos la función que estudiamos recién, $f(x,y) = x^2 + y^2$, y por otro lado la función $f_2(x,y) = \sqrt{x^2+y^2}$, podemos ver que en ambos casos todas las curvas de nivel son circunferencias (distintas circunferencias, separadas entre sí de distinta manera), pero naturalmente los gráficos de las funciones son distintos.

Otra forma de obtener información adicional es observando cortes complementarios. Por ejemplo, podemos ver cómo es el comportamiento cuando $y = 0$, es decir, estudiar $f(x,0)$. En el caso de $f(x,y) = x^2 + y^2$, resulta $f(x,0) = x^2$. Tenemos entonces que el corte de la superficie-gráfico de f con el plano $y = 0$, es una parábola. Lo mismo sucede si cortamos con el plano $x = 0$ (¡verifíquelo!). El gráfico de esta función tiene como nombre *paraboloide*, pues puede ser visto también como el resultado de girar una parábola respecto al eje vertical.

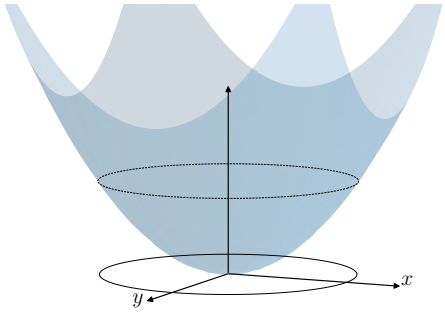


Figura 5.2: Gráfico de la función $f(x,y) = x^2 + y^2$, y una curva de nivel dibujada en el "piso" (dominio de la función), y también representada sobre el gráfico.

Sin embargo, cuando consideramos la función $f_2(x,y) = \sqrt{x^2+y^2}$, al cortar con el plano $y = 0$, resulta $f(x,0) = \sqrt{x^2+0} = |x|$, y de la misma forma, $f(0,y) = |y|$. Es decir, la intersección del gráfico de f_2 con los planos $x = 0$ e $y = 0$ tiene la forma de la gráfica del valor absoluto. Pero recordemos que las curvas de nivel son circunferencias. El gráfico de esta función es entonces un *cono*:

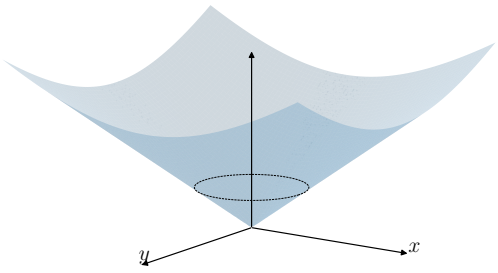


Figura 5.3: Gráfico de la función $f(x,y) = \sqrt{x^2+y^2}$, y una curva de nivel representada sobre el gráfico.

Ejercicio 5.2
Sea $f : \mathbb{R}^2 \to \mathbb{R}$ la función definida como $f(x,y) = x^2 - y^2$. Estudiar las curvas de nivel y los cortes con los planos $x = 0$ y $y = 0$, y corroborar que el gráfico de la función es como el de la Figura 5.4. Por su forma cerca del origen, se denomina usualmente silla de montar a esta función.

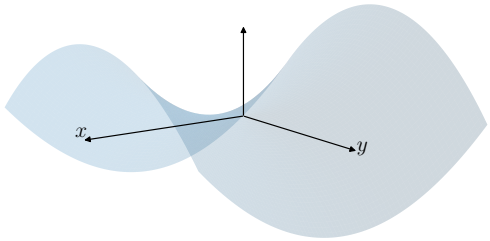


Figura 5.4: Gráfico de la función $f(x,y) = x^2 - y^2$

Ejercicio 5.3
Estudiar las curvas de nivel de la función $f_1(x,y) = |x| + |y|$. Puede ser útil recordar el ejercicio 4.4 del capítulo anterior. Estudiar también los cortes con los planos $x = 0$ e $y = 0$, y concluya cómo es la forma del gráfico de f_1.

5.2. Límites y continuidad

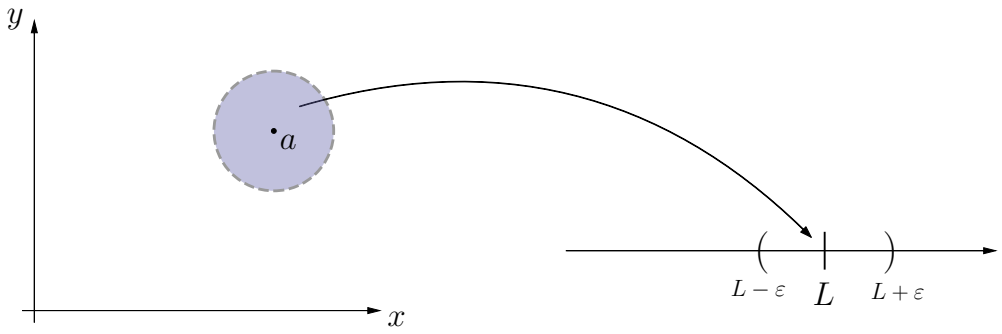
En una variable estudiamos básicamente funciones definidas en un intervalo, y límites cuando nos acercamos a un punto dentro del intervalo (o a un borde). La unidimensionalidad del dominio en estos casos, hace que haya básicamente dos "lados"; por los cuales acercarnos al punto donde estamos calculando el límite (por derecha y por izquierda, o incluso solo una de ellas, si el punto está en el borde). Cuando tenemos funciones definidas en el plano por ejemplo, nos podemos acercar a un punto por infinitas direcciones. Esto es un cambio fundamental respecto a las funciones de una variable. Veremos consecuencias de este hecho en breve.

Para poder definir el límite de una función en un punto, debemos poder acercarnos al punto mediante elementos del dominio, para poder evaluar la función allí. Esa propiedad, de poder acercarnos a un punto mediante elementos de un conjunto, ya la estudiamos en el capítulo anterior, y se llama ser *punto de acumulación* del conjunto, en este caso del dominio.

Ya hemos hecho una generalización del concepto de límite, en el capítulo anterior, cuando definimos convergencia de sucesiones en $\mathbb{R}^n$. Bastó con tener definida una distancia, para poder extender la definición en una variable. Haremos lo mismo para el límite de funciones.

Definición 5.4. Dado un conjunto $D \subset \mathbb{R}^n$, una función $f : D \to \mathbb{R}$, y $a \in \mathbb{R}^n$ un punto de acumulación de D, decimos que

$$\lim_{x \to a} f(x) = L \text{ sii } \forall \varepsilon > 0, \exists \delta > 0 \text{ tal que } \forall x \in B^*(a, \delta) \cap D \text{ se cumple } f(x) \in B(L, \varepsilon).$$



Observar primero que el concepto es el mismo que ya conocemos: para que los valores funcionales estén cerca de L (tan cerca como queramos), basta estar suficientemente cerca de a.

En segundo lugar, observar que no necesitamos que la función esté definida en el punto a, e incluso si está definida allí, el valor funcional $f(a)$ no influye en la definición.

Antes de comenzar con la operatoria de límites, veamos el concepto de continuidad, que ya sabemos que está muy relacionado con el de límite. La primera diferencia es que vamos a definir continuidad en puntos del dominio (recordar que para definir el límite no lo pedíamos).

Definición 5.5. Dado un conjunto $D \subset \mathbb{R}^n$, una función $f : D \to \mathbb{R}$, y $a \in \mathbb{R}^n$ un punto de D, decimos que f es continua en a sii

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ tal que } \forall x \in B(a, \delta) \cap D \text{ se cumple } f(x) \in B(f(a), \varepsilon).$$

Observación 5.6
El punto a debe estar en el dominio (en particular porque calculamos $f(a)$), pero no necesariamente debe ser un punto de acumulación. Así, podemos distinguir dos casos:

1. *Si a es un punto de acumulación de D, entonces la definición de continuidad coincide con la de límite, con $L = f(a)$. Es decir, en este caso,*

$$f \text{ es continua en } a \iff \lim_{x \to a} f(x) = f(a).$$

2. *Si a no es de acumulación de D, entonces es un punto aislado. Es decir, existe un radio $\delta > 0$ tal que no hay puntos de D en $B(a, \delta)$. Entonces, una función f siempre será continua en los puntos aislados del dominio*[2].

[2]Pues dado cualquier ε, basta tomar el radio δ que hace que en la bola $B(a, \delta)$ solo se encuentre a.

Ejercicio 5.7
Escribir la negación de la definición de continuidad. Es decir, completar la siguiente frase: f no es continua en a sii ...

Esta negación aparece explicada más abajo, pero de todas maneras intente razonarla y escribirla antes de llegar a ese segmento, pues es un importante ejercicio.

Veremos ahora un resultado que relaciona estos conceptos con las sucesiones. Esto además de ser sumamente útil para algunas demostraciones, también nos facilitará algunos cálculos.

Este resultado es una versión más general del Teorema 2.31 que enunciamos, a modo de adelanto, para el caso unidimensional. Enunciaremos ahora dos versiones de este resultado, una para la continuidad y otra para el límite en un punto. En vistas de la última observación, es muy fácil deducir uno a partir del otro, por lo cual demostraremos solamente una de las versiones.

Teorema 5.8. Sea un conjunto $D \subset \mathbb{R}^n$, una función $f : D \to \mathbb{R}$, y $a \in \mathbb{R}^n$ un punto de acumulación de D. Entonces

$$\lim_{x\to a} f(x) = L \iff \text{para toda sucesión } x_k \text{ de elementos de } D\setminus\{a\} \text{ tal que } \lim_{k\to\infty} x_k = a,$$
$$\text{tenemos que } \lim_{k\to\infty} f(x_k) = L.$$

Teorema 5.9. Sea un conjunto $D \subset \mathbb{R}^n$, una función $f : D \to \mathbb{R}$, y $a \in \mathbb{R}^n$ un punto de D. Entonces

$$f \text{ es continua en } a \iff \text{para toda sucesión } x_k \text{ de elementos de } D \text{ tal que } \lim_{k\to\infty} x_k = a,$$
$$\text{tenemos que } \lim_{k\to\infty} f(x_k) = f(a).$$

Demostración. ($\Longrightarrow$) Debemos demostrar que toda sucesión convergente al punto a, cumple que la sucesión de imágenes $f(x_k)$ converge a $f(a)$. Sea entonces una sucesión genérica x_k de elementos de D, tal que $x_k \to a$, y buscaremos demostrar que $f(x_k)$ converge a $f(a)$.

Ahora, por un lado sabemos que f es continua en a (es la hipótesis), por lo tanto para estar cerca de $f(a)$ basta estar suficientemente cerca de a. Por otro lado,

Como queremos demostrar que $\lim_{k\to\infty} f(x_k) = f(a)$, tomemos un $\varepsilon > 0$ genérico. Para este valor de ε, como f es continua, existe un radio $\delta > 0$ tal que todos los puntos que estén a menos de δ de a, estarán a distancia menor que ε de $f(a)$.

Pero como x_k converge al punto a, podemos encontrar un momento a partir del cual la sucesión se encuentra en $B(a,\delta)$. Es decir, para este valor de δ que nos dio la continuidad de f[3], existe un $k_0 \in \mathbb{N}$ tal que $x_k \in B(a,\delta)$ para todo $k \geqslant k_0$. Pero entonces, por cómo elegimos el δ con la continuidad de f, a partir de este k_0, tenemos que $f(x_k) \in B(f(a),\varepsilon)$, como queríamos probar.

[3]De alguna manera, estamos usando este δ como el "ε" de la definición de convergencia para sucesiones.

($\Longleftarrow$) Probaremos la continuidad de f en a por absurdo. Supongamos entonces que f no es continua en a, y buscaremos un ejemplo que contradiga la hipótesis. Esto es[4], buscamos construir una sucesión x_k convergente al punto a, pero tal que sus imágenes $f(x_k)$ no converjan a $f(a)$. Recordemos que estamos suponiendo que f no es continua en a, y vimos que esto significa que podemos encontrar puntos arbitrariamente cerca de a, cuyas imágenes estén "lejos" de $f(a)$. Formalicemos esta idea construyendo la sucesión buscada.

Como suponemos que f no es continua en a, sabemos que existe un $\varepsilon > 0$ tal que para todo $\delta > 0$, hay al menos un punto $x \in B(a,\delta)$ tal que su imagen $f(x)$ dista más que ε de $f(a)$. Dado este ε fijo (le negación de continuidad nos dice que **existe un** ε tal que ...), para cada valor de $\delta > 0$, podemos encontra (al menos un) punto que diste menos de δ de a, pero cuya imagen diste más de ε de $f(a)$. Tomemos entonces valores de δ cada vez más chicos, por ejemplo $\delta_k = 1/k$, y para cada uno de ellos, llamamos x_k a ese punto tal que $x_k \in B(a,\delta_k)$, pero $f(x_k) \notin B(f(a),\varepsilon)$.

De esta manera, construimos una sucesión x_k que converge al punto a por construcción (pues el término número k está a menos de $1/k$ de a), pero sus imágenes $f(x_k)$ están siempre a más de ese ε fijo de $f(a)$, y por lo tanto la sucesión $f(x_k)$ no puede ser convergente a $f(a)$.

Llegamos entonces a un absurdo, culminando así la demostración. □

Ejemplo 5.10
Consideremos la función $f : \mathbb{R}^2 \to \mathbb{R}$ definida como $f(x,y) = x^3y^2e^{x-y}$, e intentemos calcular el límite $\lim_{(x,y)\to(0,0)} f(x,y)$.

Hasta el momento tenemos dos herramientas para calcular este límite: la definición y el Teorema 5.8 que probamos recién. Podríamos tomar el camino de la definición, fijar un valor de ε e intentar encontrar un δ que cumpla lo buscado, pero en general este camino es tedioso.

*Tomemos entonces una sucesión genérica en el plano, que sea convergente al origen. Es decir, $(x_k,y_k) \to (0,0)$. ¿Por qué genérica? Porque para usar el Teorema 5.9 (en la dirección del recíproco más específicamente) debemos demostrar que "para **toda** sucesión tal que ...". La única manera de hacer esto es tomar una sucesión genérica, convergente al origen, y estudiar qué sucede con sus imágenes.*

Veamos entonces la sucesión (real) $f(x_k,y_k) = x_k^3y_k^2e^{x_k-y_k}$. Como la sucesión de la primera coordenada, x_k, tiende a cero, también lo hace x_k^3. De la misma forma y_k^2 tiende a cero, y por lo tanto también lo hace el producto $x_k^3y_k^2$. Por otro lado, la resta $x_k - y_k$ también tiende a cero, de donde $e^{x_k-y_k} \to 1$. En definitiva:

$$f(x_k,y_k) = \underbrace{x_k^3y_k^2}_{\to 0} \cdot \underbrace{e^{x_k-y_k}}_{\to 1} \to 0.$$

[4]La hipótesis en este caso es que para toda sucesión x_k convergente al punto a, se cumple que $f(x_k)$ tiende a $f(a)$. La negación de que **todas** las sucesiones convergentes a a cumplen eso, es que **existe una** sucesión x_k cuyo límite es a **pero** que $f(x_k)$ no converge a $f(a)$.

Como habíamos tomado una sucesión genérica, concluimos que $\lim_{(x,y)\to(0,0)} f(x,y) = 0$.

Ejemplo 5.11

Veamos otro ejemplo. Consideremos la función $f : \mathbb{R}^2 \setminus (0,0) \to \mathbb{R}$, *definida como* $f(x,y) = \frac{xy}{x^2+y^2}$, *y estudiemos el límite en el origen. Nuevamente, consideremos una sucesión genérica* $(x_k, y_k) \to (0,0)$, *y estudiemos sus imágenes:*

$$f(x_k, y_k) = \frac{x_k y_k}{x_k^2 + y_k^2}.$$

Tanto el numerador como el denominador tienden a cero cuando k tiende a infinito, y por lo tanto no podemos afirmar nada aún. En el ejemplo anterior no sucedía esto, no teníamos términos que competían entre sí.

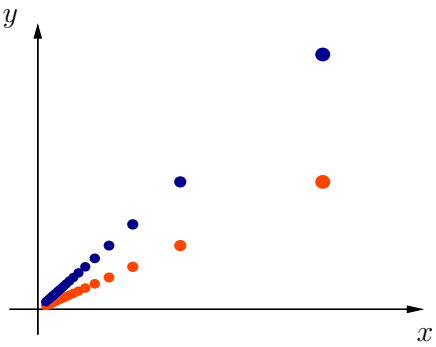
Como no podemos deducir el comportamiento en general de $f(x_k, y_k)$, *elijamos una sucesión particular y veamos qué sucede. Sea entonces* $(x_k, y_k) = (1/k, 1/k)$. *Entonces en este caso tenemos*

$$f(x_k, y_k) = \frac{\frac{1}{k}\frac{1}{k}}{\left(\frac{1}{k}\right)^2 + \left(\frac{1}{k}\right)^2} = \frac{1}{2}.$$

Concluimos que, en este caso, $\lim_{(x,y)\to(0,0)} f(x,y) = \frac{1}{2}$.

¿Este resultado depende de la elección de la sucesión? Veamos qué sucede con otra elección. Sea entonces $(x_k, y_k) = (\frac{1}{k}, \frac{1}{2k})$, *que es la otra sucesión representada en la figura. Entonces, ahora resulta*

$$f(x_k, y_k) = \frac{\frac{1}{k}\frac{1}{2k}}{\left(\frac{1}{k}\right)^2 + \left(\frac{1}{2k}\right)^2} = \frac{\frac{1}{2}}{1 + \frac{1}{4}} = \frac{2}{5}.$$



El resultado numérico no es lo importante en sí. Lo que debemos destacar es que el resultado es distinto, dependiendo de la sucesión elegida. Con esto ya podemos concluir que la función no tiene límite en el origen, pues si tuviera límite L, entonces los límites de $f(x_k, y_k)$ *deberían ser siempre L, para toda sucesión que se acerque al* $(0,0)$.

Eligiendo sucesiones particulares entonces, podemos concluir la **no** existencia de un límite, en el caso que encontremos resultados límites distintos de $f(x_k, y_k)$. Pero cuidado, porque si elegimos dos sucesiones particulares, como hicimos recién en el ejemplo, y llegamos a que $f(x_k, y_k)$ converge al mismo valor en ambos casos, no estamos en condiciones de garantizar la existencia del límite. Volveremos sobre esto enseguida.

Observemos que en el ejemplo anterior, en definitiva nos acercamos al origen con dos sucesiones, cuyo recorrido en $\mathbb{R}^2$ está incluido en rectas[5]. Lo que veremos a continuación es de alguna manera equivalente a esto, pero obviando las sucesiones.

Límites direccionales

Tomemos una función $f : \mathbb{R}^2 \to \mathbb{R}$, para la cual queremos calcular el límite en el origen. Lo que haremos es estudiar la función en un subconjunto unidimensional del dominio: veremos el comportamiento en puntos del dominio (x,y) de la forma $y = mx$ por ejemplo.

Retomemos la última función estudiada, $f(x,y) = \frac{xy}{x^2+y^2}$, y estudiemos la función de una variable que resulta de restringir el dominio a las rectas $y = mx$. Es decir:

$$\lim_{x\to 0} f(x,mx) = \lim_{x\to 0} \frac{mx^2}{x^2+mx^2} = \frac{m}{1+m^2}.$$

Por un lado, recuperamos (para $m = 1$ y $m = 1/2$) los resultados que obtuvimos con sucesiones. Por otro lado, la observación importante es que este límite **depende de** m, es decir, depende de cuál es la dirección por la cual nos acercamos. Por lo tanto, concluimos que esta función no tiene límite en el origen (a lo que ya habíamos llegado con las dos sucesiones).

Observemos que haciendo variar la pendiente de la recta m entre todos los reales, cubrimos todas las rectas que pasan por el origen salvo la vertical. Para esto, debemos estudiar la función restringiendo el dominio a los puntos con $x = 0$, es decir, estudiar $f(0,y)$. De todas maneras, dado que la mayor utilidad de los limites direccionales se da para concluir la no existencia de límites, en general no es necesario estudiar tan finamente todas las direcciones.

Veamos otro ejemplo.

Ejemplo 5.12
Consideremos la función $f : \mathbb{R}^2 \setminus (0,0) \to \mathbb{R}$, definida como $f(x,y) = \frac{xy^2}{x^2+y^4}$, y estudiemos los límites direccionales.

$$\lim_{x\to 0} f(x,mx) = \lim_{x\to 0} \frac{m^2x^3}{x^2+m^4x^4} = \lim_{x\to 0} \frac{m^2x^3}{x^2(1++m^4x^2)} = 0.$$

Es fácil ver que si estudiamos $f(0,y)$ también obtenemos límite cero. Tenemos entonces que los límites direccionales, por cualquier recta, son cero. Sin embargo, esto no alcanza para concluir que existe el límite. Lo único que podemos afirmar es que, si el límite existe, debe ser cero.

[5] Aunque perfectamente podríamos haber elegido $(x_k,y_k) = (\frac{1}{k}, \frac{1}{k^2})$, y en ese caso el recorrido quedaría incluido en una parábola.

Veamos qué sucede cuando nos acercamos, no por una recta, sino por una parábola conveniente, $x = y^2$.

$$\lim_{y\to 0} f(y^2, y) = \lim_{y\to 0} \frac{y^2 y^2}{y^4 + y^4} = \frac{1}{2}.$$

Encontramos una forma de acercarnos al origen, de manera que los valores funcionales tienden a $1/2$. Por lo tanto, aunque todos los límites direccionales dieron el mismo resultado, el límite no existe.

Observación 5.13
Es importante recordar que los límites direccionales ***no permiten garantizar*** *la existencia del límite. Son útiles para hallar un candidato a límite, o para demostrar que el límite no existe.*

Coordenadas polares

En los límites anteriores, con $(x.y)$ tendiendo al $(0,0)$, diferenciamos dos elementos: que el punto (x,y) se acerque al origen (en términos de distancia), y cómo lo hace, en términos de "dirección".

En el Capítulo 1 ya nos encontramos con estos conceptos cuando definimos la notación polar de un complejo. En efecto, describimos a un punto en el plano complejo mediante su módulo (distancia al origen) y el ángulo que forma con el eje horizontal.

Tomemos esta misma descripción para los puntos del plano. Es decir, podemos escribir un punto $(x,y) \in \mathbb{R}^2$ mediante dos números, ρ y θ, de la siguiente forma:

$$\begin{aligned} x &= \rho\cos(\theta), \\ y &= \rho\sin(\theta). \end{aligned}$$

Recordemos que como ρ es la distancia al origen debe ser un número positivo. Por otro lado, alcanza que el ángulo θ pueda recorrer intervalo $[0, 2\pi)$ para describir cualquier dirección en el plano. Más adelante, en el Capítulo 7, estudiaremos con más detalle esta representación. Con lo que vimos hasta el momento es suficiente para usarlo como herramienta para calcular límites.

Veamos con un ejemplo cómo es el procedimiento.

Ejemplo 5.14
Consideremos el límite $\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^2+y^2}$. *La función escrita con las variables ρ y θ es:*

$$\frac{x^2y}{x^2+y^2} \quad\rightsquigarrow\quad \frac{(\rho\cos(\theta))^2\rho\sin(\theta)}{(\rho\cos(\theta))^2+(\rho\sin(\theta))^2} = \frac{\rho^3\cos^2(\theta)\sin(\theta)}{\rho^2(\cos(\theta)^2+\sin(\theta)^2)} = \rho\cos^2(\theta)\sin(\theta).$$

Observemos que, como el límite que queremos calcular es con (x,y) tendiendo al origen, con las nuevas variables debemos calcular el límite con ρ tendiendo a cero.

Entonces tenemos que calcular:

$$\lim_{\rho\to 0} \rho\cos^2(\theta)\sin(\theta) = 0.$$

La observación importante aquí es que el factor $\cos^2(\theta)\sin(\theta)$ está acotado (entre -1 y 1 por ejemplo), y por lo tanto tenemos el límite de un término acotado, por otro que tiende a cero.

El argumento aquí es el siguiente. Tomemos un $\varepsilon > 0$ cualquiera, y sea M una cota del valor absoluto del segundo factor. Como el primer término tiende a cero, podemos encontrar un valor δ a partir del cual ese término es más chico que ε/M. En definitiva, para los valores de $\rho < \delta$, el valor de f es menor que ε. Y la condición $\rho < \delta$ es exactamente que (x,y) esté en una bola de centro $(0,0)$ y radio δ, que es lo que aparece en la definición de límite.

Veamos otro ejemplo.

Ejemplo 5.15
Consideremos el límite cuando (x,y) tiende al $(0,0)$ de la función

$$f(x,y) = \begin{cases} \frac{xy}{x+y} & si\ x+y \neq 0 \\ 0 & si\ x+y = 0. \end{cases}$$

Dejemos de lado por un segundo la dirección donde la función vale cero por definición. En el resto del dominio, tenemos que el límite en coordenadas polares es

$$\lim_{\rho\to 0} \frac{\rho^2\cos(\theta)\sin(\theta)}{\rho(\cos(\theta)+\sin(\theta))} = \lim_{\rho\to 0} \rho\cdot\frac{\cos(\theta)\sin(\theta)}{\cos(\theta)+\sin(\theta)}.$$

Estamos en una situación similar que en el ejemplo anterior. Un factor que depende solo de ρ, y que tiende a cero, y otro factor que depende solo de θ:

$$\lim_{\rho\to 0} \rho\cdot\frac{\cos(\theta)\sin(\theta)}{\cos(\theta)+\sin(\theta)}$$

*Podríamos estar tentados a argumentar de la misma forma (algo que tiende a cero por algo acotado), y llegar a que el límite debe ser cero. Sin embargo, hay que prestar especial atención en este caso, pues el segundo término **no está acotado**. En efecto, para ángulos cercanos a $-\pi/4$ por ejemplo, podemos hacer el denominador tan chico como queramos.*

Se puede ver que en realidad la función en cuestión no tiene límite cuando (x,y) tiende al origen.

Observación 5.16
Si fijamos un valor de θ, y hacemos tender ρ a cero, nos estamos acercando al origen por una dirección dada. Esto es equivalente a los límites direccionales que vimos más arriba. Sin embargo, este procedimiento de fijar un valor de θ no es lo que hacemos en general cuando calculamos límites por polares.

Ejercicio 5.17
Utilizar el cambio a polares para decidir la existencia de los siguientes límites, y calcularlos si corresponde.

$$\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2+y^2} \qquad \lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2+y^2}} \qquad \lim_{(x,y)\to(0,0)} \frac{x^3+y^3}{x^2+y^2}.$$

Volviendo a la continuidad y sus propiedades, veamos que el Teorema 5.9 permite demostrar los resultados clásicos de operaciones con funciones continuas, de una manera muy simple.

Proposición 5.18. Si f y g son dos funciones continuas en a, entonces:

1. $f+g$ es continua en a.
2. $f \cdot g$ es continua en a.
3. Si $g(a) \neq 0$, entonces f/g es continua en a.

Demostración. Demostremos solamente la continuidad de la suma, para ejemplificar la metodología. Utilizaremos el Teorema 5.9 en las dos direcciones: como **sabemos** que f y g son continuas, usaremos la dirección del directo para ellas; y como **queremos demostrar** que $f+g$ es continua, lo haremos utilizando el la dirección del recíproco. Tomemos entonces una sucesión genérica x_k en $\mathbb{R}^n$, que sea convergente al punto a. Como f es continua en a, entonces $\lim_{k\to\infty} f(x_k) = f(a)$. De la misma forma, como g es continua en a, $\lim_{k\to\infty} g(x_k) = g(a)$.

Por lo tanto, como sabemos que el límite de la suma de sucesiones es la suma de los límites (ver Teorema 2.11), resulta que

$$\lim_{k\to\infty} (f+g)(x_k) = \lim_{k\to\infty} f(x_k) + g(x_k) = f(a) + g(a).$$

Ahora, como es cierto que para toda sucesión x_k que tiende al punto a, sus imágenes por la función $f+g$ tienden a $(f+g)(a)$, en virtud de la dirección del recíproco del Teorema 5.9, concluimos que $f+g$ es continua. □

Las demostraciones de las otras propiedades son similares, así como la continuidad de la composición, que enunciaremos en lo que sigue, y cuya prueba también queda como ejercicio.

Proposición 5.19. Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función continua en $a \in \mathbb{R}^n$, y $g : \mathbb{R} \to \mathbb{R}$ una función continua en $f(a) \in \mathbb{R}$. Entonces la composición $g \circ f$ es continua en a.

5.3. Teoremas para funciones continuas

Cuando pensamos en resultados importantes sobre las funciones continuas en una variable que hayamos visto en un primer curso de cálculo, aparecen varios naturalmente: teoremas de valor medio (Bolzano o Darboux), Teorema de Weierstrass sobre la existencia de extremos en compactos, Teorema de Cantor sobre la continuidad uniforme en compactos, y Teorema Fundamental del Cálculo.

Para algunos de estos resultados resulta difícil pensar en generalizaciones. Por ejemplo, no tenemos aún una noción de integral en dimensiones mayores, por lo que no podemos imaginar un Teorema Fundamental. Sí definimos conjuntos compactos, por lo que veremos a continuación el Teorema de Weierstrass:

Teorema 5.20 (Weierstrass)**.** Sea $C \subset \mathbb{R}^n$ un conjunto compacto, y $f : C \to \mathbb{R}$ una función continua. Entonces f alcanza mínimo y máximo en C. Es decir, existen $x_m \in C$ y $x_M \in C$, tales que $f(x) \geqslant f(x_m)$ para todo $x \in C$, y $f(x) \leqslant f(x_M)$ para todo $x \in C$.

Demostración. Probaremos primero que f es acotada (superiormente). Supongamos entonces que no lo es, es decir, que para todo $B \in \mathbb{R}$, existe un punto x en C tal que $f(x) > B$. Tomemos entonces una sucesión creciente de valores para B, por ejemplo $B_k = k$. Entonces, para cada uno de estos B_k, existe un punto en C, que llamaremos x_k, tal que $f(x_k) > B_k$, y por lo tanto $\lim_{k\to\infty} f(x_k) = \infty$. Construimos de esta manera una sucesión x_k, que por estar contenida en un conjunto compacto, tiene una subsucesión convergente. Llamemos x_{k_j} a esta subsucesión, y $\bar{x} \in C$ a su límite.

Como f es continua en todo el conjunto C (en particular en el punto $\bar{x}$), tenemos que para cualquier sucesión que tienda a $\bar{x}$, sus imágenes por f tienden a $f(\bar{x})$. Sin embargo, la subsucesión construida x_{k_j} tiende a $\bar{x}$, pero sus imágenes tienden a infinito. Llegamos a una contradicción, que surge de suponer la no acotación de f, por lo que concluimos que f es acotada.

La segunda parte de la demostración es muy similar, por lo que la dejaremos como un ejercicio guiado: Sea M el supremo del conjunto $\{f(x) : x \in C\}$ (¿Por qué existe?). Utilizando la propiedad del supremo, con la cual podemos encontrar puntos del conjunto tan cerca de M como queramos, construir una sucesión x_k de puntos de C, tal que sus imágenes $f(x_k)$ converjan a M. Utilizar la compacidad de C para encontrar una subsucesión convergente de x_k, y finalmente utilizar la continuidad de f en ese límite de la subsucesión para probar que $f(L) = M$. □

El Teorema que vimos recién tiene tres ingredientes fundamentales: la continuidad, que el dominio C sea cerrado, y que el dominio sea acotado. ¿Son los tres necesarios?

Ejercicio 5.21
No es necesario ir a varias variables. Se pueden pensar ejemplos con funciones de una variable.

1. *Busque un ejemplo de una función $f : C \to \mathbb{R}$, con C compacto (pero f no necesariamente continua), donde la función no alcance máximo y/o mínimo.*

2. *Busque un ejemplo de una función $f : C \to \mathbb{R}$ continua, con C cerrado (pero no necesariamente acotado), donde la función no alcance máximo y/o mínimo.*

3. *Busque un ejemplo de una función $f : C \to \mathbb{R}$ continua, con C acotado (pero no necesariamente cerrado), donde la función no alcance máximo y/o mínimo.*

También es cierto que una función continua en un compacto es uniformemente continua, pero no haremos énfasis en esta propiedad.

Además...

En el curso de cálculo en una variable, la mayoría de los teoremas estaban enunciados para funciones continuas definidas en un intervalo $[a,b]$, pero, ¿qué es lo importante del dominio para cada teorema? Para el Teorema de Weierstrass ya vimos que lo fundamental del dominio es su compacidad. Pasa lo mismo para el Teorema de Cantor sobre continuidad uniforme. Sin embargo, para los teoremas de valor medio (pensemos en Bolzano por ejemplo), que también están enunciados en intervalos $[a,b]$, la compacidad no parece ser la responsable, sino más bien el poder "llegar de un punto al otro del dominio". Es decir, si $f(a)$ es negativo y $f(b)$ es positivo, en algún lugar del camino entre a y b, tiene que haber un cero. Esto sucede siempre en un intervalo, pero no en la unión de dos intervalos disjuntos, por ejemplo.
Un resultado local[a] que sí se generaliza inmediatamente es el que sigue.

Proposición 5.22. Sea $f:\mathbb{R}^n \to \mathbb{R}$, continua en $a \in \mathbb{R}^n$, y tal que $f(a) > 0$. Entonces existe un entorno $B(a,\delta)$ tal que f es positiva en todo el entorno.

Demostración. Basta considerar $\varepsilon = \frac{f(a)}{2}$ en la definición de continuidad. El δ correspondiente cumple lo enunciado. □

Tenemos entonces que si f es positiva en un punto, entonces hay un entorno del punto donde también es positiva. En definitiva, el conjunto de los puntos donde f es positiva, $\{x \in \mathbb{R}^n : f(x) > 0\}$, es un conjunto abierto. Lo mismo sucede con el conjunto de puntos donde f es estrictamente negativa.
Supongamos entonces que tenemos un conjunto C, y una función continua $f : C \to \mathbb{R}$, que alcance valores positivos y valores negativos, pero que no alcance el cero (de alguna manera sería una negación de la tesis de Bolzano). Entonces, tendríamos que el dominio se puede particionar en dos conjuntos: el conjunto de puntos con imagen positiva, y el conjunto de puntos con imagen negativa. Estos dos conjuntos son abiertos (lo vimos recién), y disjuntos (claramente). Entonces escribimos el dominio como unión de dos conjuntos abiertos y disjuntos. Aquellos conjuntos que **no se pueden** particionar de esta forma, como unión de dos abiertos disjuntos, se llaman conjuntos *conexos*. La intuición que nos da este nombre es bastante acertada: son conjuntos formados "por una sola pieza".
Lo que probamos entonces es una versión del Teorema de Bolzano: si $C \in \mathbb{R}^n$ es un conjunto conexo, $f : C \to \mathbb{R}$ es continua, y hay dos puntos a y b en C tales que $f(a) > 0$ y $f(b) < 0$, entonces existe $c \in C$ tal que $f(c) = 0$.
Esta última discusión, sobre la generalización de teoremas de valor medio, no es de gran importancia para lo que resta del curso, pero la presentamos aquí porque resulta divertida, y un buen ejercicio de razonamiento, cuestionamiento, y generalización.

[a] En el sentido de comportamientos locales, alrededor de un punto.

Capítulo 6

Diferenciabilidad

6.1. Derivadas parciales y direccionales

En el capítulo anterior vimos que el concepto de continuidad (o de límite) presentaba ciertas similitudes y diferencias con la noción que teníamos para funciones de una variable. La definición es exactamente una generalización de la continuidad conocida, donde utilizamos las nociones de bola para generalizar los entornos de la forma $(a-\delta, a+\delta)$ por ejemplo. Es decir, habiendo comprendido las nociones básicas de topología en $\mathbb{R}^n$ (y en $\mathbb{R}$ en particular), esta definición de continuidad no ofrece mayores dificultades. Sin embargo, desde el punto de vista operativo, para hacer cálculos de límites, encontramos una diferencia mayor.

En esta primera parte de derivabilidad sucederá lo contrario de alguna manera: será sencillo hacer cuentas, pero las definiciones requieren un proceso más cuidadoso, aunque de aspecto sean muy similares a las definiciones de derivada que acostumbramos utilizar.

Cuando pensamos en la derivada de una función de una variable, una de las primeras cosas que se nos vienen a la mente es la de crecimiento de la función. Esto está bien definido para funciones con dominio en $\mathbb{R}$, pero ¿podemos definir crecimiento de una función definida en el plano? Si estamos parados en un punto $a \in \mathbb{R}^2$, ¿qué quiere decir que la función crece? ¿Cuando nos movemos hacia dónde?. Esta diferencia en la dimensión del dominio ya nos causó problemas para el cálculo de límites, y en ese momento estudiamos la restricción de la función a rectas. Recordemos que estos limites direccionales nos daban alguna información parcial, pero no garantizaban la existencia del límite. Tomemos este camino también para las primeras nociones de derivabilidad.

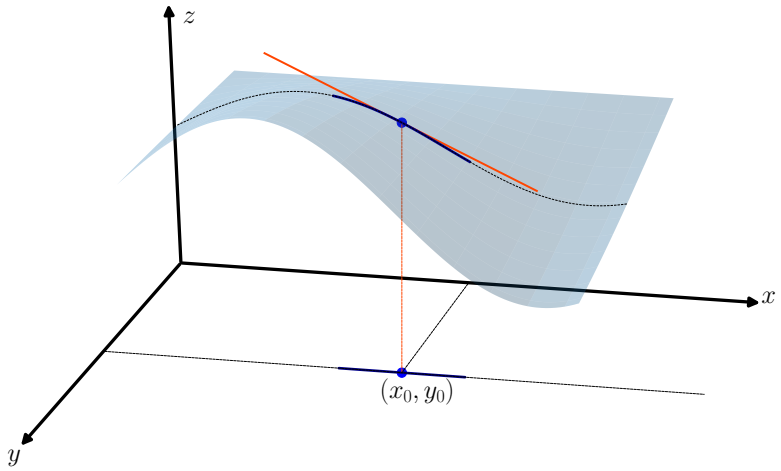
Consideremos la función $f(x,y) = x^2 + y^2$, y miremos lo que resulta cuando nos restringimos a los puntos del plano con coordenada $y = 0$, es decir, $f(x,0)$. Esto corresponde, si se quiere, a cortar el gráfico de f en $\mathbb{R}^3$ con un plano vertical. La curva que resulta de la

intersección de la superficie (que es un paraboloide como vimos) con el plano, es el gráfico de la función de una variable x^2, que conocemos bien. Para esta función de una variable sí podemos hablar de derivada, y por ejemplo la derivada en el origen es nula. Si volvemos a la función f definida en el plano, esto lo interpretamos así: si estamos parados en el $(0,0)$ y nos movemos en la dirección del eje x, la función f "sale con pendiente horizontal". Este número cero, que resultó de derivar x^2 en el origen, nos da información sobre el comportamiento de f en una dirección. Formalicemos esto.

Definición 6.1. Sea $f:\mathbb{R}^2\to\mathbb{R}$, y (x_0,y_0) un punto de $\mathbb{R}^2$. Definimos la derivada parcial de f respecto a la variable x en el punto (x_0,y_0) como el siguiente límite (si existe):

$$\frac{\partial f}{\partial x}(x_0,y_0)=\lim_{h\to 0}\frac{f(x_0+h,y_0)-f(x_0,y_0)}{h}.$$

Observar que el límite de la definición es un límite de una variable (h es un real que tiende a cero), que es exactamente el límite del cociente incremental de la función de una variable $f(x,y_0)$, que resulta de fijar la variable y al valor y_0.



De manera análoga se define la derivada parcial respecto a la variable y.

Ejemplo 6.2
Veamos que ambas derivadas parciales de $f(x,y)=x^2+y^2$ en el origen son nulas.

$$\frac{\partial f}{\partial x}(0,0)=\lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h}=\lim_{h\to 0}\frac{h^2-0}{h}=0.$$

$$\frac{\partial f}{\partial y}(0,0)=\lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h}=\lim_{h\to 0}\frac{h^2-0}{h}=0.$$

Para la motivación de la definición, estudiamos la función restringida a uno de los ejes. Si denominamos explícitamente a esta función como g, es decir, fijamos la segunda coordenada al valor $y = y_0$, y consideramos la función de una variable $g(x) = f(x, y_0)$. Entonces la definición de la derivada de g en x_0 coincide con la definición de la derivada parcial de f respecto a x en el punto (x_0, y_0).

$$g'(x_0) = \lim_{h\to 0} \frac{g(x_0+h) - g(x_0)}{h} = \lim_{h\to 0} \frac{f(x_0+h, y_0) - g(x_0, y_0)}{h} = \frac{\partial f}{\partial x}(x_0, y_0).$$

Esta observación nos permite hacer cálculos de manera sencilla: si queremos hallar la derivada parcial respecto de x, basta considerar la variable y como constante, y operar como sabemos hacerlo.

Ejemplo 6.3
Si queremos hallar $\frac{\partial f}{\partial x}(1,1)$ *para la función* $f(x,y) = x^2y + 5y$*, debemos pensar en la función* $x^2 \cdot 1 + 5$*, derivarla respecto a x, y evaluar en* $x = 1$*. Resulta entonces* $\frac{\partial f}{\partial x}(1,1) = 2$*.*

La definición se generaliza trivialmente para funciones definidas en $\mathbb{R}^n$.

Se utilizan varias notaciones para la derivada parcial, por ejemplo la que vimos $\frac{\partial f}{\partial x}$, o simplemente $\partial_x f$ (o incluso indicando con el subíndice la coordenada respecto a la que estamos derivando: $\partial_1 f$). Otra forma muy usada es f_x. En este texto utilizaremos $\frac{\partial f}{\partial x}$ o f_x.

Hasta ahora trabajamos con las derivadas parciales en un punto. Así como hicimos con las funciones de una variable, podemos definir por ejemplo la función *derivada parcial respecto a* x, que a cada punto del plano le asigna el valor de la derivada parcial en ese punto. Es decir, en este caso tenemos $\frac{\partial f}{\partial x} : \mathbb{R}^2 \to \mathbb{R}$.

Ejercitemos estos nuevos conceptos con algunos ejemplos.

Ejemplos 6.4

1. *Si* $f(x,y) = x^2y + ye^{2x}$*, entonces tenemos que:*

$$\frac{\partial f}{\partial x}(x,y) = 2xy + 2ye^{2x}.$$
$$\frac{\partial f}{\partial y}(x,y) = x^2 + e^{2x}.$$

2. *Tomemos ahora* $f(x,y) = (y-x)^2 + x^2y^2$*. Entonces:*

$$\frac{\partial f}{\partial x}(x,y) = -2(y-x) + 2xy^2.$$
$$\frac{\partial f}{\partial y}(x,y) = 2(y-x) + x^2 2y.$$

Como vemos en este último ejemplo, desde el punto de vista de calcular derivadas no tenemos mayores diferencias con lo que conocemos. Cabe preguntarnos entonces si sigue siendo cierto que la derivabilidad implica continuidad. En este caso, la pregunta que debemos hacernos es: si existen las derivadas parciales, ¿necesariamente la función es continua?.

En el capítulo anterior nos acercábamos por diferentes rectas para calcular límites direccionales, y concluimos que esta información no era suficiente para garantizar la existencia del límite. Ahora, con las derivadas parciales, estamos teniendo en cuenta lo que sucede solamente en dos direcciones, por lo que todo parece indicar que esto no será suficiente para decidir sobre la continuidad, como lo muestra el siguiente ejemplo.

Ejemplo 6.5
Consideremos la función definida como

$$f(x,y) = \begin{cases} 1 & \text{si } x = 0 \text{ o } y = 0 \\ 0 & \text{en otro caso.} \end{cases}.$$

Entonces, como la función es constante sobre los ejes, las derivadas parciales en el origen existen y valen 0*. Sin embargo, la función no es continua en* $(0,0)$*, ya que no existe* $\lim_{(x,y)\to(0,0)} f(x,y)$.

El análisis anterior nos lleva inevitablemente a tomar en cuenta el comportamiento de la función en otras direcciones. El procedimiento será análogo: restringimos la función a una recta, y estudiamos la función de una variable que resulta.

Definición 6.6. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ una función, (x_0, y_0) un punto de $\mathbb{R}^2$, y un vector dirección $v = (v_1, v_2)$. Entonces definimos la derivada direccional de f respecto a la dirección v en el punto (x_0, y_0) como el siguiente límite, si existe:

$$\frac{\partial f}{\partial v}(x_0,y_0) = \lim_{h\to 0} \frac{f(x_0+hv_1, y_0+hv_2) - f(x_0,y_0)}{h}.$$

Es decir, es exactamente el límite del cociente incremental de la función que resulta de restrigir f a la recta de dirección v por el punto (x_0, y_0).

Observar que las derivadas parciales son casos particulares de las derivadas direccionales, cuando el vector de dirección v es alguno de los vectores canónicos e_i. Por ejemplo, si $v = (1,0)$, obtenemos la derivada parcial respecto a x.

Ejemplo 6.7
Calculemos la derivada direccional respecto a la dirección $v = (1,1)$ *de la función* $f(x,y) = x^2 + y^2$*. Entonces:*

$$\frac{\partial f}{\partial v}(x_0,y_0) = \lim_{h\to 0} \frac{f(x_0+hv_1, y_0+hv_2) - f(x_0,y_0)}{h} = \lim_{h\to 0} \frac{(x_0+h)^2 + (y_0+h)^2 - x_0^2 - y_0^2}{h}$$

$$= \lim_{h\to 0} \frac{(x_0^2 + 2x_0h + h^2) + (y_0^2 + 2y_0h + h^2) - x_0^2 - y_0^2}{h}$$

$$= \lim_{h\to 0} \frac{2x_0h + h^2 + 2y_0h + h^2}{h} = 2x_0 + 2y_0.$$

Ejercicio 6.8 (Un Teorema de valor medio)
Sea $f : \mathbb{R}^2 \to \mathbb{R}$ una función, $a \in \mathbb{R}^2$ un punto, y v una dirección, tal que la derivada direccional $\frac{\partial f}{\partial v}(x,y)$ existe para todo (x,y) en el segmento $[a, a+v]$. Demostrar que existe un $\theta \in (0,1)$ tal que $f(a+v) - f(a) = \frac{\partial f}{\partial v}(a+\theta v)$. Sugerencia: considerar la función $\alpha : [0,1] \to \mathbb{R}^2$ definida como $\alpha(t) = a + tv$, y aplicar el teorema de valor medio de Lagrange a la función $f \circ \alpha$.

Volviendo al ejemplo 6.5, es fácil ver que si tomamos cualquier dirección v distinta de las canónicas, la derivada direccional en el origen no existe[1].

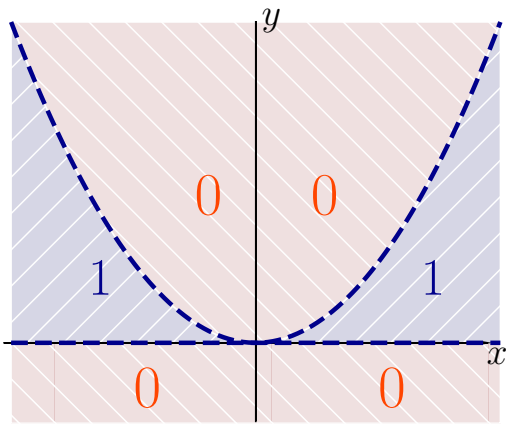
Ahora podemos tener en cuenta el comportamiento de la función en todas las direcciones. Es momento de preguntarse nuevamente, entonces, si la existencia de todas las derivadas direccionales implica la continuidad. Veamos un ejemplo en esta dirección.

Ejemplo 6.9

Consideremos la función definida como

$$f(x,y) = \begin{cases} 1 & \text{si } 0 < y < x^2 \\ 0 & \text{en otro caso,} \end{cases}$$

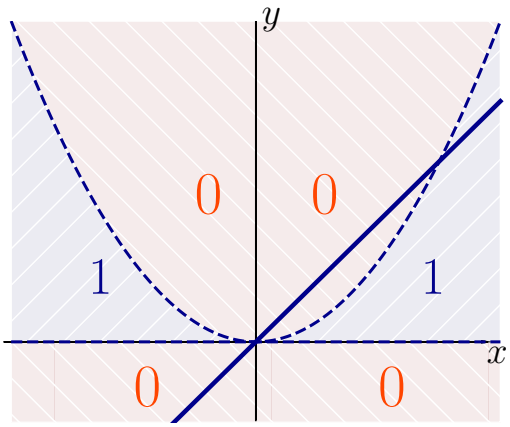
y estudiemos, por un lado, si existen todas las derivadas direccionales en el origen, y por otro lado la continuidad en $(0,0)$.
Como se ve en la figura, la función toma el valor cero en todo el semiplano inferior y por arriba de la parábola $y = x^2$, y toma el valor uno entre el eje horizontal y dicha parábola.



Para estudiar las derivadas direccionales, fijemos una dirección $v = (v_1, v_2)$, y veamos la derivada direccional según esta dirección. Es decir, siguiendo la definición, debemos calcular $\frac{\partial f}{\partial v}(0,0) = \lim_{h\to 0} \frac{f(0+hv_1, 0+hv_2) - f(0,0)}{h}$.

[1] Pues $f((0,0)+hv) = 0$ para todo h distinto de cero, pero $f(0,0) = 1$.

Por ejemplo, como en los ejes la función es constante cero, entonces las derivadas parciales existen y son nulas.



Tomemos ahora la dirección $v = (1,1)$. Entonces debemos calcular $\lim_{h\to 0} \frac{f(h,h)-f(0,0)}{h}$. Ahora bien, $f(h,h)$ será cero o uno, dependiendo de si para ese h caemos en la región pintada o no. En la figura se puede ver que los puntos de la recta cerca del origen (para valores de h chicos), caen dentro de la zona donde la función vale cero. Por lo tanto el límite $\lim_{h\to 0} \frac{f(h,h)-f(0,0)}{h}$ vale cero, ya que el numerador es nulo para valores de h en un cierto intervalo de cero.

Esta situación se repite para cualquier vector $v = (v_1, v_2)$. En otras palabras, cualquier recta que pasa por el $(0,0)$, para valores cercanos al origen, cae dentro de la zona donde f vale cero. Esto se puede ver, por ejemplo, calculando los puntos de intersección de una recta cualquiera con la parábola (uno de ellos será siempre el origen).

Concluimos entonces que todas las derivadas direccionales existen y son cero: $\frac{\partial f}{\partial v}(0,0) = 0$.

Estudiemos ahora la continuidad. En cualquier entorno del origen hay puntos donde la función vale cero, y puntos donde la función vale uno, por lo que la función no puede ser continua en $(0,0)$. Otra forma de verlo es encontrando una curva tal que, cuando calculamos el límite restringiéndonos a esa curva, el resultado es distinto a $f(0,0)$. Por ejemplo, en todos los puntos de la curva $y = x^2/2$, tenemos que f vale 1, por lo tanto el límite es 1, mientras que $f(0,0) = 0$.

En definitiva, esta es un ejemplo de una función con todas las derivadas direccionales en el origen, pero no es continua en ese punto. La existencia de derivadas direccionales no nos dice nada sobre la continuidad.

El siguiente ejercicio constituye otro ejemplo de comportamiento similar.

Ejercicio 6.10
Considere la función definida como

$$f(x,y) = \begin{cases} \frac{x^3 y}{x^6+y^2} & si\ (x,y) \neq (0,0) \\ 0 & si\ (x,y) = (0,0) \end{cases}.$$

1. *Probar que existen todas las derivadas direccionales en el origen.*
2. *Probar que f no es continua en $(0,0)$. Sugerencia: estudiar $f(x,x^3)$ y $f(x,0)$.*

Es claro, por otro lado, que la continuidad no puede implicar la existencia de derivadas parciales o direccionales. En primer lugar, porque esto ya era falso en funciones de una variable, pero además porque es fácil construir ejemplos de funciones continuas que no tengan derivadas parciales. Por ejemplo $f(x,y) = |x| + |y|$ en el origen (Ejercicio: verificar que f es continua en $(0,0)$, y que no existen las derivadas parciales en ese punto).

La continuidad y la existencia de derivadas direccionales son entonces conceptos independientes. Por lo tanto, estamos ante una noción de derivabilidad que no generaliza bien todas las ideas que traemos del cálculo en una variable.

6.2. Diferenciabilidad

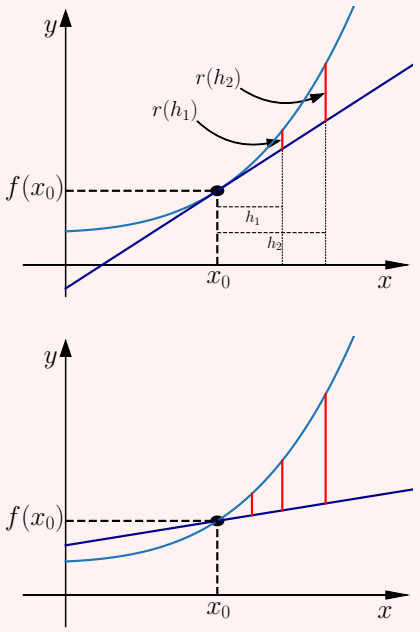
Para definir derivadas parciales, nos restringimos a rectas y consideramos el límite del cociente incremental, tal cual lo hacíamos con funciones definidas en $\mathbb{R}$. Ahora trataremos de generalizar las nociones de una variable, pero mirando otras propiedades que teníamos.

Específicamente, sea $f : \mathbb{R} \to \mathbb{R}$ una función derivable, y x_0 un punto de $\mathbb{R}$. Entonces, tenemos que

$$f(x_0+h) = f(x_0) + f'(x_0)h + r(h) \tag{6.1}$$

con $r : \mathbb{R} \to \mathbb{R}$ una función tal que $\lim_{h\to 0} \frac{r(h)}{h} = 0$, es decir, que tiende a cero más rápido que h. Es decir, localmente, la función se aproxima bien por un polinomio de primer grado. Geométricamente, en este caso corresponde a la aproximación por la recta tangente.

Aprovechemos que ya estamos familiarizados con las funciones de una variable, para observar la importancia de la propiedad del resto.
La función $r(h)$ mide cuán lejos se encuentra la aproximación lineal de la función, como se ilustra en la figura de arriba, con la recta tangente.
Ahora, siempre que la aproximación lineal pase por el punto $(x_0, f(x_0))$, el error $r(h)$ tiende a cero cuando h tiende a cero, ¡sin importar si la pendiente de la recta tiene algo que ver con el gráfico de la función! En la figura de abajo podemos ver este hecho con una aproximación cualquiera.
Es decir, pedir que $r(h) \to 0$ con h, no es pedir casi nada. Para que la aproximación sea buena, el error tiene que tender a cero **rápido**, en este caso más rápido que h.



Tomemos esta idea para generalizarla a funciones de varias variables, pensando en funciones de $\mathbb{R}^2$ a $\mathbb{R}$ para fijar ideas. Entonces, la idea geométrica de aproximar el gráfico de la función por una recta, pasa ahora a ser una aproximación por un plano (o por un hiperplano en dimensiones mayores). Analíticamente, estas estructuras (rectas, planos) se corresponden con transformaciones lineales. En efecto, en la aproximación (6.1), $f'(x_0)h$ es una transformación lineal de $\mathbb{R}$ a $\mathbb{R}$ (es decir, multiplicar por una constante, que en este caso es la derivada en x_0), aplicada al incremento h.

Definamos entonces diferenciabilidad exactamente así:

Definición 6.11. Sean $f : \mathbb{R}^n \to \mathbb{R}$, y $a \in \mathbb{R}^n$. Decimos que la función f es diferenciable en a sii existe una transformación lineal $df_a : \mathbb{R}^n \to \mathbb{R}$ tal que

$$f(a+h) = f(a) + df_a(h) + r(h)$$

con $r : \mathbb{R}^n \to \mathbb{R}$ una función que cumple $\lim_{h\to 0} \frac{r(h)}{||h||} = 0$.

Observar que h es ahora un incremento en $\mathbb{R}^n$. Es decir, nos movemos de a hacia $a+h$, y la diferencia del valor funcional, $f(a+h)-f(a)$, se puede aproximar bien por una transformación lineal evaluada en el incremento $df_a(h)$, y lo que sobra o falta, $r(h)$, tiende a cero rápido.

Aprovechemos la intuición geométrica que estamos obteniendo para las funciones definidas en $\mathbb{R}^2$, y veamos cómo se traduce esta definición para este caso particular. Para eso, recordemos que las transformaciones lineales de $\mathbb{R}^2$ a $\mathbb{R}$ tienen la forma $T(x,y) = Ax + By$. La definición de diferenciabilidad entonces la podemos escribir así:

Definición 6.12. Sean $f : \mathbb{R}^2 \to \mathbb{R}$, y $a = (x_0, y_0) \in \mathbb{R}^2$. Decimos que la función f es diferenciable en (x_0, y_0) sii existen dos reales A y B tales que

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + A\Delta x + B\Delta y + r(\Delta x, \Delta y)$$

con $r : \mathbb{R}^2 \to \mathbb{R}$ una función que cumple $\lim_{(\Delta x,\Delta y)\to(0,0)} \frac{r(\Delta x,\Delta y)}{||(\Delta x,\Delta y)||} = 0$.

Aquí el incremento que antes llamamos h, ahora es $(\Delta x, \Delta y)$, y la transformación lineal aplicada al incremento es $A\Delta x + B\Delta y$.

Al igual que observamos para el caso de una variable, la condición del resto es fundamental: Hay infinitos valores de A y B que hacen que el resto tienda a cero con h, la clave de una buena aproximación lineal es que el resto tienda a cero más rápido que la norma de h.

Veamos que esta definición de diferenciabilidad sí garantiza la continuidad, y también estudiemos cómo se relaciona con las nociones de derivabilidad que vimos.

Pensemos en la definición de diferenciabilidad en $\mathbb{R}$, que es en definitva lo escrito en (6.1). Esencialmente, dice que la función f puede aproximarse bien por un polinomio de primer

grado. Ahora, esto es un caso particular del Teorema de Taylor, que afirma además que este polinomio (el que aproxima bien), es único, y es necesariamente el que tiene como coeficiente a la derivada de f evaluada en el punto. Para el caso general, donde tenemos varios coeficientes, serán las derivadas parciales las que jueguen ese papel.

Teorema 6.13. Sea $f : \mathbb{R}^2 \to \mathbb{R}$ una función diferenciable en un punto (x_0, y_0). Entonces:

1. f es continua en (x_0, y_0).
2. Existen las derivadas parciales en (x_0, y_0) y valen $\frac{\partial f}{\partial x}(x_0, y_0) = A$ y $\frac{\partial f}{\partial y}(x_0, y_0) = B$.
3. Existen todas las derivadas direccionales, y si $v = (v_1, v_2)$, entonces $\frac{\partial f}{\partial v}(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)v_1 + \frac{\partial f}{\partial y}(x_0, y_0)v_2$.

Demostración. 1. De la definición de diferenciabilidad, tenemos que

$$f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0) + A\Delta x + B\Delta y + r(\Delta x, \Delta y).$$

Entonces,

$$\begin{aligned}\lim_{(\Delta x,\Delta y)\to(0,0)} f(x_0 + \Delta x, y_0 + \Delta y) &= \lim_{(\Delta x,\Delta y)\to(0,0)} f(x_0, y_0) + A\Delta x + B\Delta y + r(\Delta x, \Delta y) \\ &= f(x_0, y_0) + \lim_{(\Delta x,\Delta y)\to(0,0)} A\Delta x + B\Delta y + r(\Delta x, \Delta y) \\ &= f(x_0, y_0).\end{aligned}$$

donde usamos que $r(\Delta x, \Delta y)$ tiende a cero[2] cuando $(\Delta x, \Delta y) \to (0, 0)$.

En definitva, $\lim_{(\Delta x,\Delta y)\to(0,0)} f(x_0 + \Delta x, y_0 + \Delta y) = f(x_0, y_0)$, que es exactamente la definición de continuidad en (x_0, y_0) (ya que es un punto interior al dominio).

2. Planteemos la definición de derivada parcial, y utilicemos la diferenciabilidad para calcular el límite:

$$\begin{aligned}\frac{\partial f}{\partial x}(x_0, y_0) &= \lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \\ &= \lim_{h\to 0} \frac{f(x_0, y_0) + A \cdot h + B \cdot 0 + r(h, 0) - f(x_0, y_0)}{h} \\ &= A + \lim_{h\to 0} \frac{r(h, 0)}{h} = A.\end{aligned}$$

en la última igualdad, usamos que $\lim_{h\to 0} \left|\frac{r(h,0)}{h}\right| = 0$. La derivada parcial respecto a y se calcula de la misma forma.

[2] Observar que sabemos más que eso: tenemos que $\lim_{(\Delta x,\Delta y)\to(0,0)} \frac{r(\Delta x,\Delta y)}{||(\Delta x,\Delta y)||} = 0$, por lo tanto, en particular, $\lim_{(\Delta x,\Delta y)\to(0,0)} r(\Delta x, \Delta y) = 0$

3. Procedemos de manera similar para calcular las derivadas direccionales, pero ahora sabiendo que los coeficientes A y B de la definición de diferenciabilidad son las derivadas parciales, que ahora notaremos $f_x(x_0,y_0)$ y $f_y(x_0,y_0)$ por comodidad. Sea $v=(v_1,v_2)$:

$$\begin{aligned}\frac{\partial f}{\partial v}(x_0,y_0) &= \lim_{h\to 0}\frac{f(x_0+hv_1,y_0+hv_2)-f(x_0,y_0)}{h}\\ &= \lim_{h\to 0}\frac{f(x_0,y_0)+f_x(x_0,y_0)\cdot hv_1+f_y(x_0,y_0)\cdot hv_2+r(hv_1,hv_2)-f(x_0,y_0)}{h}\\ &= f_x(x_0,y_0)\cdot v_1+f_y(x_0,y_0)\cdot v_2+\lim_{h\to 0}\frac{r(hv_1,hv_2)}{h}.\end{aligned}$$

Observemos que en el límite que nos resta calcular, no tenemos exactamente la función resto dividido la norma del incremento, ya que en lugar de $||(hv_1,hv_2)||$ tenemos solamente h. Esto lo podemos solucionar multiplicando y dividiendo entre la norma de v, para lograr obtener en el denominador la norma del incremento:

$$\lim_{h\to 0}\left|\frac{r(hv_1,hv_2)}{h}\right| = \lim_{h\to 0}\left|\frac{r(hv_1,hv_2)}{h||(v_1,v_2)||}\right| ||(v_1,v_2)|| = \lim_{h\to 0}\left|\frac{r(hv_1,hv_2)}{||(hv_1,hv_2)||}\right| ||(v_1,v_2)|| = 0.$$

□

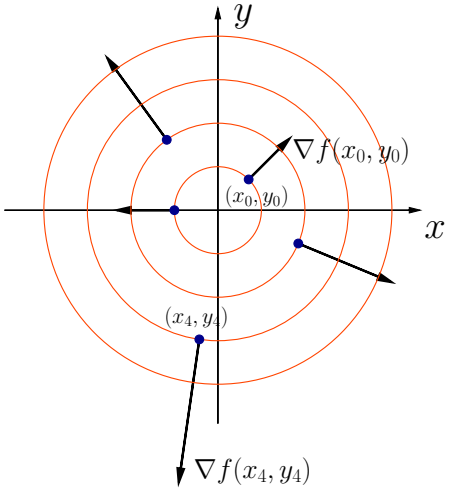
En principio, para una función cualquiera, las derivadas direccionales (incluídas las derivadas parciales) pueden ser completamente independientes entre sí. Lo que nos dice la parte tres del último teorema es que cuando f es diferenciable, nos alcanza con conocer sus derivadas parciales para tener todas las derivadas direccionales. Esto es consistente con la idea del plano que aproxima al gráfico de la función: conociendo el plano (que en definitiva es conocer las derivadas parciales), tenemos información sobre el comportamiento local de la función en todas las direcciones.

Esta información queda codificada en el gradiente, que es un vector formado por las derivadas parciales:

Definición 6.14. Dada $f:\mathbb{R}^2\to\mathbb{R}$ diferenciable en (x_0,y_0), definimos el vector gradiente de f en (x_0,y_0) como $\nabla f(x_0,y_0)=\left(\frac{\partial f}{\partial x}(x_0,y_0),\frac{\partial f}{\partial y}(x_0,y_0)\right)$.

Observación: Con esta notación, podemos expresar la derivada direccional como el producto interno del gradiente con la dirección: $\frac{\partial f}{\partial v}(x_0,y_0)=\langle\nabla f(x_0,y_0),v\rangle$.

Consideremos ahora un vector v de norma uno, y las derivadas direccionales de f respecto a v. Como tenemos que $\frac{\partial f}{\partial v}(x_0,y_0) = \langle \nabla f(x_0,y_0), v\rangle$, entonces la derivada direccional se maximiza cuando los vectores v y $\nabla f(x_0,y_0)$ son colineales. Es decir, la dirección del gradiente es la dirección de máximo crecimiento de f. Además, como veremos más adelante, el gradiente es siempre perpendicular a las curvas de nivel.
En la figura podemos observar algunas curvas de nivel de una función, y los vectores gradientes en algunos puntos.



En la definición 6.11, el dominio de la función es todo $\mathbb{R}^n$, pero en realidad lo que necesitamos es que el punto a, donde definimos la diferenciabilidad, sea un punto interior al dominio.

Diremos que una función es diferenciable cuando lo es en todos los puntos de su dominio.

Ejemplos 6.15
Estudiemos, utilizando la definición, la diferenciabilidad de algunas funciones. El Teorema 6.13 nos dice cuál es la transformación lineal que deberíamos probar: la que tiene como coeficientes las derivadas parciales en el punto.

1. *Consideremos la función $f(x,y) = 2x+3y+4$, y estudiemos la diferenciabilidad en el origen. La derivada parcial respecto a x es $f_x(x,y) = 2$ en todo punto, y por otro lado $f_y(x,y) = 3$. Por lo tanto, estos son los coeficientes de la transformación lineal a utilizar en la definición de diferenciabilidad. Es decir, para chequear si f es diferenciable en el origen, debemos estudiar si es cierto que*

$$f(0+\Delta x, 0+\Delta y) = f(0,0) + 2\Delta x + 3\Delta y + r(\Delta x, \Delta y),$$

con el resto tendiendo a cero más rápido que la norma del incremento. Pero $f(0,0) = 4$, por lo tanto, despejando de la expresión anterior, obtenemos que el resto es la función nula, y por lo tanto cumple con la condición deseada. En realidad, ya sabíamos que esto debería suceder: la función f es lineal (y su gráfico es un plano), por lo tanto la mejor aproximación lineal es ella misma, y entonces el resto es cero.

2. *Consideremos ahora la función* $f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & si\ (x,y) \neq (0,0) \\ 0 & si\ (x,y) = (0,0). \end{cases}$

Para estudiar la diferenciabilidad en el origen, debemos calcular las derivadas parciales en ese punto. La función f es constante en los ejes, así que tendremos por ejemplo

$$f_x(0,0) = \lim_{h\to 0} \frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0} \frac{0-0}{h} = 0.$$

De igual manera llegamos a que la derivada respecto a y se anula, y por lo tanto $\nabla f(0,0) = (0,0)$. Además, $f(0,0) = 0$, y por lo tanto todos los términos de la aproximación lineal son nulos. Es decir, de la definición de diferenciabilidad de f en el origen, el resto es exactamente la función:

$$f(0+\Delta x, 0+\Delta y) = 0 + 0\cdot\Delta x + 0\cdot\Delta y + r(\Delta x,\Delta y).$$

Para ver si f es diferenciable entonces, debemos estudiar el límite del resto sobre la norma del incremento. Si dicho límite resulta ser cero, la función es diferenciable, y si resulta algo distinto de cero, o no existe, la función no es diferenciable.

$$\begin{aligned}\lim_{(\Delta x,\Delta y)\to(0,0)} \frac{r(\Delta x,\Delta y)}{\sqrt{\Delta x^2+\Delta y^2}} &= \lim_{(\Delta x,\Delta y)\to(0,0)} \frac{1}{\sqrt{\Delta x^2+\Delta y^2}} \frac{\Delta x\Delta y}{\sqrt{\Delta x^2+\Delta y^2}} \\ &= \lim_{(\Delta x,\Delta y)\to(0,0)} \frac{\Delta x\Delta y}{\Delta x^2+\Delta y^2}.\end{aligned}$$

Este límite no existe (es exactamente el límite del ejemplo 5.11), y por lo tanto la función no es diferenciable.

3. *Modifiquemos un poco la función del ejemplo anterior. Estudiemos la diferenciabilidad en el origen de* $f(x,y) = \begin{cases} \frac{x^2y}{\sqrt{x^2+y^2}} & si\ (x,y)\neq(0,0) \\ 0 & si\ (x,y)=(0,0). \end{cases}$

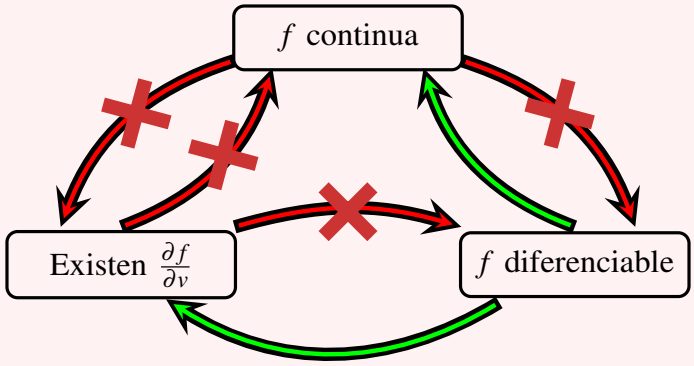
Como nuevamente tenemos que f es constante en los ejes, el gradiente es el vector nulo $\nabla f(0,0) = (0,0)$. También $f(0,0) = 0$, por lo que nuevamente el resto es exactamente la función. Tenemos entonces que:

$$\lim_{(\Delta x,\Delta y)\to(0,0)} \frac{r(\Delta x,\Delta y)}{\sqrt{\Delta x^2+\Delta y^2}} = \lim_{(\Delta x,\Delta y)\to(0,0)} \frac{\Delta x^2\Delta y}{\Delta x^2+\Delta y^2}.$$

Este último límite es efectivamente cero (ver capítulo anterior), por lo que esta función sí es diferenciable en el origen.

Implicancias

A modo de resumen, visualicemos en un diagrama las implicancias entre tres conceptos fundamentales: la continuidad (arriba), la existencia de todas las derivadas direccionales (abajo a la izquierda), y la diferenciabilidad (abajo a la derecha).
Las flechas verdes corresponden a resultados que tenemos, y que implican por ejemplo que si una función es diferenciable entonces es continua.
Las flechas rojas significan que la implicancia señalada no es cierta (por ejemplo, que hay funciones continuas que no son diferenciables).



Ejercicio 6.16
Buscar contraejemplos (en estas notas por ejemplo) para cada una de las flechas rojas.

Plano Tangente

Volviendo a la idea geométrica que nos sirvió de intuición para definir diferenciabilidad, veamos explícitamente cuál es la ecuación del plano tangente en un punto al gráfico de una función diferenciable.

Sea entonces una función f diferenciable en el punto (x_0, y_0). Entonces la ecuación del plano tangente por (x_0, y_0) es:

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Es decir, es el gráfico de la parte lineal (sin el resto) de la aproximación usada en la definición de diferenciabilidad.

Por ejemplo, si f es la función $f(x,y) = x^2 + y^2$, y consideramos el punto $(x_0, y_0) = (1,2)$, entonces la ecuación del plano tangente por el $(1,2)$ es

$$z = 5 + 2(x-1) + 4(y-2).$$

Hasta el momento, para comprobar la diferenciabilidad de una función, solamente teníamos la definición. Por lo tanto, debíamos calcular el límite del resto sobre la norma del incremento.

Veamos un resultado que nos permitirá asegurar la diferenciabilidad de una función más fácilmente. La demostración se vuelve engorrosa por momentos, debido a la notación, por lo que recomendamos leerla con paciencia, e intentar no perder de vista el objetivo (que en este caso será comprobar la propiedad del resto).

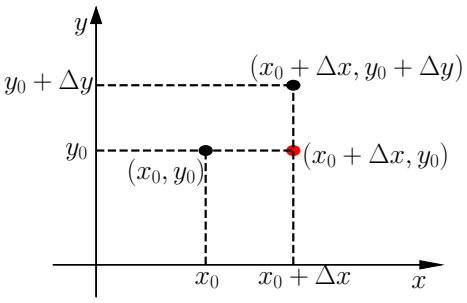
Teorema 6.17 (Condición suficiente de diferenciabilidad)**.** Sea $f:\mathbb{R}^2\to\mathbb{R}$ una función y $(x_0,y_0)\in\mathbb{R}^2$ un punto, tales que las derivadas parciales de f existen en una bola de centro (x_0,y_0) y ambas son continuas en (x_0,y_0). Entonces f es diferenciable en (x_0,y_0).

Demostración. Debemos ir a la definición de diferenciabilidad, y demostrar que el resto tiende a cero más rápido que la norma del incremento $(\Delta x,\Delta y)$. Despejando el resto, tenemos

$$r(\Delta x,\Delta y)=f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)-f_x(x_0,y_0)\Delta x-f_y(x_0,y_0)\Delta y.$$

Luego nos ocuparemos de dividir entre la norma del incremento.

Ahora, como tenemos información sobre las derivadas parciales, debemos buscar la forma de tener incrementos con alguna variable fija, que es donde podemos aplicar resultados con las derivadas parciales.
Entonces, para ir de (x_0,y_0) a $(x_0+\Delta x,y_0+\Delta y)$, lo haremos en dos tramos, como indica la figura. Analíticamente, lo que haremos es sumar y restar f evaluada en el punto $(x_0+\Delta x,y_0)$.



Es decir, escribimos

$$f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)=f(x_0+\Delta x,y_0+\Delta y)-f(x_0+\Delta x,y_0)+f(x_0+\Delta x,y_0)-f(x_0,y_0).$$

Observemos que en los dos primeros términos, nos movimos solamente en la segunda variable, y en los últimos dos términos, solamente en la primera variable. Específicamente, en los últimos dos términos, la segunda variable es y_0. Podemos considerar entonces la función de una variable que resulta de fijar y_0. Llamémosle $\varphi(x)=f(x,y_0)$, y tenemos en particular que su derivada vale $\varphi'(x)=f_x(x,y_0)$.

Podemos escribir entonces, por el Teorema de Valor Medio para derivadas (Lagrange), que

$$f(x_0+\Delta x,y_0)-f(x_0,y_0)=\varphi(x_0+\Delta x)-\varphi(x_0)=\varphi'(x_0+\theta\Delta x)\Delta x=f_x(x_0+\theta\Delta x,y_0)\Delta x,$$

con $\theta\in(0,1)$. De igual manera, si llamamos $\psi(y)=f(x_0+\Delta x,y)$, tenemos que

$$f(x_0+\Delta x,y_0+\Delta y)-f(x_0+\Delta x,y_0)=\psi(y_0+\Delta y)-\psi(y_0)=f_y(x_0+\Delta x,y_0+\tilde\theta\Delta y)\Delta y,$$

con $\tilde\theta\in(0,1)$, nuevamente aplicando Lagrange. Volviendo, tenemos hasta ahora que

$$f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)=\underbrace{f(x_0+\Delta x,y_0+\Delta y)-f(x_0+\Delta x,y_0)}_{f_y(x_0+\Delta x,y_0+\tilde\theta\Delta y)\Delta y}+\underbrace{f(x_0+\Delta x,y_0)-f(x_0,y_0)}_{f_x(x_0+\theta\Delta x,y_0)\Delta x},$$

y por lo tanto, sustituyendo en la expresión para el resto:

$$r(\Delta x,\Delta y) = f_x(x_0+\theta\Delta x,y_0)\Delta x + f_y(x_0+\Delta x,y_0+\tilde{\theta}\Delta y)\Delta y - f_x(x_0,y_0)\Delta x - f_y(x_0,y_0)\Delta y.$$

Ahora sí, dividimos entre la norma del incremento, y tomamos factor común Δx y Δy:

$$\frac{r(\Delta x,\Delta y)}{||(\Delta x,\Delta y)||} = \frac{\Delta x\big[f_x(x_0+\theta\Delta x,y_0)-f_x(x_0,y_0)\big]+\Delta y\big[f_y(x_0+\Delta x,y_0+\tilde{\theta}\Delta y)-f_y(x_0,y_0)\big]}{||(\Delta x,\Delta y)||}$$
$$= \frac{\Delta x}{||(\Delta x,\Delta y)||}\big[f_x(x_0+\theta\Delta x,y_0)-f_x(x_0,y_0)\big] + \frac{\Delta y}{||(\Delta x,\Delta y)||}\big[f_y(x_0+\Delta x,y_0+\tilde{\theta}\Delta y)-f_y(x_0,y_0)\big].$$

Ahora, el término $\frac{\Delta x}{||(\Delta x,\Delta y)||}$ está acotado, así como $\frac{\Delta y}{||(\Delta x,\Delta y)||}$. Y por otro lado, como las derivadas parciales son continuas en (x_0,y_0), cuando $(\Delta x,\Delta y)$ tiende al $(0,0)$, tenemos que $f_x(x_0+\theta\Delta x,y_0)$ tiende a $f_x(x_0,y_0)$. Llegamos entonces a que

$$\underbrace{\frac{\Delta x}{||(\Delta x,\Delta y)||}}_{Acotado}\underbrace{\big[f_x(x_0+\theta\Delta x,y_0)-f_x(x_0,y_0)\big]}_{\to 0} + \underbrace{\frac{\Delta y}{||(\Delta x,\Delta y)||}}_{Acotado}\underbrace{\big[f_y(x_0+\Delta x,y_0+\tilde{\theta}\Delta y)-f_y(x_0,y_0)\big]}_{\to 0}.$$

Por lo que concluimos que $\lim_{(\Delta x,\Delta y)\to(0,0)}\frac{r(\Delta x,\Delta y)}{||(\Delta x,\Delta y)||}=0$, y por lo tanto f es diferenciable en (x_0,y_0). □

Ejercicio 6.18
Buscar dónde se utilizó la continuidad de las derivadas en la prueba anterior.

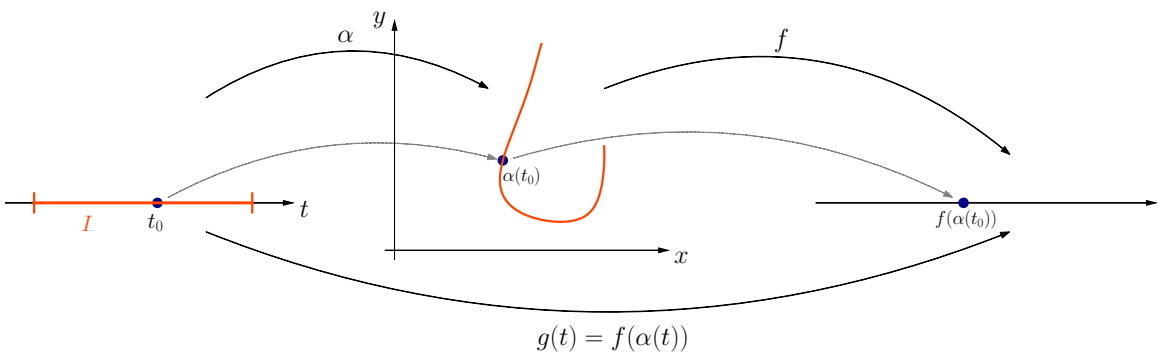
Además de este resultado, las operaciones usuales entre funciones diferenciables también resulta diferenciable, lo que permite concluir la diferenciabilidad de una gran familia de funciones. Específicamente, si $f:\mathbb{R}^2\to\mathbb{R}$ y $g:\mathbb{R}^2\to\mathbb{R}$ son diferenciables, entonces las funciones λf, $f+g$,y fg son diferenciables. Además, si g no se anula en un entorno de (x_0,y_0), también f/g es diferenciable en ese punto.

Luego de estas operaciones, lo natural sería continuar con la diferenciabilidad de la composición. Ahora, hasta el momento solamente hemos trabajado con funciones de $\mathbb{R}^2$ (o $\mathbb{R}^n$) a $\mathbb{R}$, y por lo tanto no podemos componerlas entre sí. Veamos entonces brevemente algunos elementos de funciones con codominio $\mathbb{R}^2$, para agarrar intuición geométrica.

Sea $I\subset\mathbb{R}$ un intervalo. A una función $\alpha:I\to\mathbb{R}^2$ la llamamos *curva*. Una función de un intervalo al plano, tiene dos componentes: para cada valor de $t\in I$, tenemos $\alpha(t)=(x(t),y(t))$, donde $x(t)$ e $y(t)$ son dos funciones a valores reales, que son las coordenadas en el plano del punto $\alpha(t)$.

Ejemplo 6.19
Si $\alpha:[0,2\pi]\to\mathbb{R}^2$ es $\alpha(t)=(\cos(t),\sin(t))$, obtenemos una curva que recorre la circunferencia S^1.

Podemos componer entonces una función $f:\mathbb{R}^2\to\mathbb{R}$ con una curva α, resultando en una función $g(t)=f\big(\alpha(t)\big)$, que en definitiva es una función de una variable, que toma valores reales. Es decir, una función de las que estudiamos en cálculo en una variable. ¿Qué podemos decir sobre la derivabilidad de esta función?



Teorema 6.20 (Regla de la Cadena I)**.** Sean $f:\mathbb{R}^2\to\mathbb{R}$ una función diferenciable en (x_0,y_0), y $\alpha:I\to\mathbb{R}^2$, $\alpha(t)=(x(t),y(t))$, tal que $\alpha(t_0)=(x_0,y_0)$, y las dos funciones coordenadas $x(t)$ e $y(t)$ son derivables en t_0. Entonces la composición $g(t)=(f\circ\alpha)(t)=f\big(\alpha(t)\big)$ es derivable en t_0, y su derivada vale

$$g'(t_0)=f_x(x_0,y_0)x'(t_0)+f_y(x_0,y_0)y'(t_0).$$

Demostración. Comencemos escribiendo la forma en la que usaremos la derivabilidad de las funciones $x(t)$ e $y(t)$, que no es otra que la expresión de diferenciabilidad, con la que motivamos el comienzo de esta sección.

Tenemos entonces que, como $x(t)$ es derivable en t_0, podemos escribir

$$x(t_0+h)=x(t_0)+x'(t_0)h+r_x(h),$$

donde $\frac{r_x(h)}{h}\to 0$ cuando h tiende a cero. De igual manera, tenemos que

$$y(t_0+h)=y(t_0)+y'(t_0)h+r_y(h),$$

donde $\frac{r_y(h)}{h}\to 0$ cuando h tiende a cero.

Como queremos estudiar la derivabilidad de la función g en t_0, plantearemos la definición, y estudiemos el límite del cociente incremental.

$$\begin{aligned}g'(t_0)&=\lim_{h\to 0}\frac{g(t_0+h)-g(t_0)}{h}=\lim_{h\to 0}\frac{f\big(\alpha(t_0+h)\big)-f\big(\alpha(t_0)\big)}{h}\\&=\lim_{h\to 0}\frac{f\big(x(t_0+h),y(t_0+h)\big)-f\big(x_0,y_0\big)}{h}.\end{aligned}$$

Ahora, como las funciones coordenadas de $\alpha(t)$ son diferenciables, expresaremos $x(t_0+h)$ e $y(t_0+h)$ con su aproximación de primer orden:

$$\begin{aligned} g'(t_0) &= \lim_{h\to 0} \frac{f\big(x(t_0)+x'(t_0)h+r_x(h), y(t_0)+y'(t_0)h+r_y(h)\big) - f\big(x_0,y_0\big)}{h} \\ &= \lim_{h\to 0} \frac{f\big(x_0+x'(t_0)h+r_x(h), y_0+y'(t_0)h+r_y(h)\big) - f\big(x_0,y_0\big)}{h}. \end{aligned}$$

Como tenemos la función f evaluada en (x_0,y_0) más un incremento, estamos en condiciones de usar la diferenciabilidad de f en ese punto. En este caso, el incremento es

$$(\Delta x, \Delta y) = \big(x'(t_0)h+r_x(h), y'(t_0)h+r_y(h)\big).$$

Tenemos entonces:

$$\begin{aligned} g'(t_0) &= \lim_{h\to 0} \frac{\cancel{f(x_0,y_0)} + f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + r(\Delta x,\Delta y) - \cancel{f\big(x_0,y_0\big)}}{h} \\ &= \lim_{h\to 0} \frac{f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + r(\Delta x,\Delta y)\big)}{h} \\ &= \lim_{h\to 0} \frac{f_x(x_0,y_0)\big(x'(t_0)h+r_x(h)\big) + f_y(x_0,y_0)\big(y'(t_0)h+r_y(h)\big) + r(\Delta x,\Delta y)\big)}{h} \\ &= f_x(x_0,y_0)x'(t_0) + f_y(x_0,y_0)y'(t_0) + \lim_{h\to 0} \frac{f_x(x_0,y_0)r_x(h) + f_y(x_0,y_0)r_y(h) + r(\Delta x,\Delta y)\big)}{h}. \end{aligned}$$

Resta probar que este último límite es cero.

Como $\frac{r_x(h)}{h}$ y $\frac{r_y(h)}{h}$ tienden a cero con h, los dos primeros sumandos (que se componen de estos términos multiplicados por constantes), tienden a cero.

Finalmente, cuando h tiende a cero, tenemos que $(\Delta x,\Delta y)$ tiende al $(0,0)$. Por lo tanto,

$$\lim_{h\to 0} \frac{r(\Delta x,\Delta y)}{h} = \lim_{h\to 0} \frac{r(\Delta x,\Delta y)}{\|(\Delta x,\Delta y)\|} \frac{\|(\Delta x,\Delta y)\|}{h}.$$

Sabemos que $\frac{r(\Delta x,\Delta y)}{\|(\Delta x,\Delta y)\|}$ tiende a cero, pues es la propiedad que tenemos de la diferenciabilidad de f. Veamos que $\frac{\|(\Delta x,\Delta y)\|}{h}$ está acotado. En efecto:

$$\left|\frac{\|(\Delta x,\Delta y)\|}{h}\right| = \left\|\left(\frac{\Delta x}{h},\frac{\Delta y}{h}\right)\right\| = \left\|\left(\frac{x'(t_0)h+r_x(h)}{h},\frac{y'(t_0)h+r_y(h)}{h}\right)\right\| \xrightarrow[h\to 0]{} \big\|\big(x'(t_0),y'(t_0)\big)\big\|,$$

por lo que en particular está acotado, y esto culmina la demostración. □

Interpretemos esta expresión que hallamos para la derivada de la composición. Cuando nos movemos un poco, de t_0 a t_0+h, esto genera que el punto $\alpha(t)$, que se encontraba en

(x_0, y_0) para el tiempo t_0, se mueva un poco, y como consecuencia el valor funcional, que estará evaluado en ese punto, también cambiará. ¿Cuánto? Bueno, la cantidad $x'(t_0)$ cuantiza cuánto se mueve la primera coordenada de $\alpha(t)$, que se encontraba en x_0, y por otro lado la derivada parcial respecto de x nos dice, justamente, cuánto varía f cuando movemos un poco la primera coordenada. En definitiva, el efecto del movimiento de la primera componente de $\alpha(t_0)$ (es decir, cuánto se mueve "hacia la derecha" el punto $\alpha(t)$ cuando nos movemos de t_0 a $t_0 + h$), es $f_x(x_0, y_0)x'(t_0)$. De igual manera, el movimiento vertical afecta el valor funcional, y estas componentes se suman para llegar a la expresión que acabamos de demostrar:

$$g'(t_0) = f_x(x_0, y_0)x'(t_0) + f_y(x_0, y_0)y'(y_0).$$

Observemos, además, que esta expresión la podemos re-escribir como:

$$g'(t_0) = f_x(x_0, y_0)x'(t_0) + f_y(x_0, y_0)y'(y_0) = \langle \nabla f(x_0, y_0), \alpha'(t_0) \rangle.$$

donde $\alpha'(t) = (x'(t), y'(t))$ es el vector de derivadas, que es tangente a la curva.

Utilicemos esta expresión para demostrar una propiedad importante sobre el gradiente, que ya observamos cuando los representamos gráficamente.

Supongamos que $f : \mathbb{R}^2 \to \mathbb{R}$ es una función diferenciable, y $\alpha : I \to \mathbb{R}^2$ una curva tal que f es constante en la curva. Es decir, $f\big(\alpha(t)\big) = k$, para todo t, o lo que es lo mismo, la imagen de la curva está incluida en un conjunto de nivel $C_k = \{(x, y) \in \mathbb{R}^2 : f(x, y) = k\}$. Ahora bien, como la función compuesta $g = (f \circ \alpha)$ es constante, su derivada es nula. Si por ejemplo $\alpha(t_0) = (x_0, y_0)$, , tenemos, en virtud de la regla de la cadena:

$$g'(t_0) = \langle \nabla f\big(\alpha(t_0)\big), \alpha'(t_0) \rangle = \langle \nabla f\big(x_0, y_0\big), \alpha'(t_0) \rangle = 0.$$

Es decir, el gradiente es perpendicular a las curvas de nivel.

Veamos ahora cómo utilizar la regla de la cadena, en la versión que demostramos en 6.20, para hallar las derivadas de otra composiciones. En particular, consideraremos una función $g : \mathbb{R}^2 \to \mathbb{R}^2$. Más adelante profundizaremos más sobre estas funciones que toman valores vectoriales, por ahora basta notar que una función así, tiene dos componentes (una para cada coordenada del espacio de llegada $\mathbb{R}^2$). Es decir, la función g es $g(x, y) = \big(g_1(x, y), g_2(x, y)\big)$, donde tanto g_1 como g_2 son funciones de $\mathbb{R}^2$ a $\mathbb{R}$, como las que venimos estudiando.

Teorema 6.21 (Regla de la Cadena II)**.** Sean $g : \mathbb{R}^2 \to \mathbb{R}^2$, de forma $g(x, y) = \big(g_1(x, y), g_2(x, y)\big)$, y $f : \mathbb{R}^2 \to \mathbb{R}$, que escribiremos como $f(u, v)$. Supongamos que $g(x_0, y_0) = (u_0, v_0)$, que f es diferenciable en (u_0, v_0), y que tanto g_1 como g_2 son diferenciables en (x_0, y_0). Entonces la función $h = (f \circ g)$ tiene derivadas parciales que verifican:

$$\frac{\partial h}{\partial x}(x_0, y_0) = \frac{\partial f}{\partial u}(u_0, v_0)\frac{\partial g_1}{\partial x}(x_0, y_0) + \frac{\partial f}{\partial v}(u_0, v_0)\frac{\partial g_2}{\partial x}(x_0, y_0)$$
$$\frac{\partial h}{\partial y}(x_0, y_0) = \frac{\partial f}{\partial u}(u_0, v_0)\frac{\partial g_1}{\partial y}(x_0, y_0) + \frac{\partial f}{\partial v}(u_0, v_0)\frac{\partial g_2}{\partial y}(x_0, y_0)$$

Observar que en cada una de estas dos derivadas parciales, consideramos una variable fija. Pensemos por ejemplo en $\frac{\partial h}{\partial x}$. Aquí, dejamos a la variable y como constante y_0, y por lo tanto la función $g(x,y_0) = \big(g_1(x,y_0), g_2(x,y_0)\big)$, es una función de $\mathbb{R}$ a $\mathbb{R}^2$, de las mismas características que la función α en las páginas anteriores. Para calcular entonces $\frac{\partial h}{\partial x}$, estamos utilizando exactamente el resultado de la regla de la cadena que demostramos anteriormente.

Prestar especial atención a los puntos donde están evaluadas las derivadas. Las derivadas parciales de f están evaluadas en (u_0,v_0), y las derivadas de las funciones g_1 y g_2 están evaluadas en (x_0,y_0), que son los únicos puntos donde tiene sentido evaluar cada una de estas funciones.

6.3. Derivadas de orden superior

Sea $f : \mathbb{R}^2 \to \mathbb{R}$ una función diferenciable. Las derivadas parciales son en sí mismas, también funciones definidas para cada punto del plano. Es decir, tenemos

$$f_x : \mathbb{R}^2 \to \mathbb{R} \quad \text{y} \quad f_y : \mathbb{R}^2 \to \mathbb{R},$$

y tiene sentido entonces estudiar la derivabilidad de estas funciones.

Definamos entonces las derivadas de segundo orden. Por ejemplo, si tomamos la función f_x, y hallamos su derivada respecto a x en el punto a, entonces obtenemos la derivada de f respecto de x dos veces, evaluada en a. Esto es:

$$\frac{\partial^2 f}{\partial x \partial x}(a) = \frac{\partial \left(\frac{\partial f}{\partial x}\right)}{\partial x}(a)$$

De igual manera se definen las tres derivadas restantes, que también las denotamos de distintas formas:

$$\frac{\partial^2 f}{\partial x \partial x}(a) = \frac{\partial^2 f}{\partial x^2}(a) = f_{xx}(a) \qquad \frac{\partial^2 f}{\partial x \partial y}(a) = f_{xy}(a)$$
$$\frac{\partial^2 f}{\partial y \partial y}(a) = \frac{\partial^2 f}{\partial y^2}(a) = f_{yy}(a) \qquad \frac{\partial^2 f}{\partial y \partial x}(a) = f_{yx}(a).$$

Así obtenemos las derivadas de segundo orden en un punto, pero las funciones derivadas segundas se definen de la misma forma que antes.

Veamos con un ejemplo que las operaciones siguen siendo las naturales.

Ejemplo 6.22
Consideremos la función de dos variables $f(x,y) = x\sin(y) + x^2y^2$, y calculemos sus derivadas. Tenemos:

$$f_x(x,y) = \sin(y) + 2xy^2, \qquad f_y(x,y) = x\cos(y) + 2x^2y.$$

Luego, las derivadas de segundo orden son:

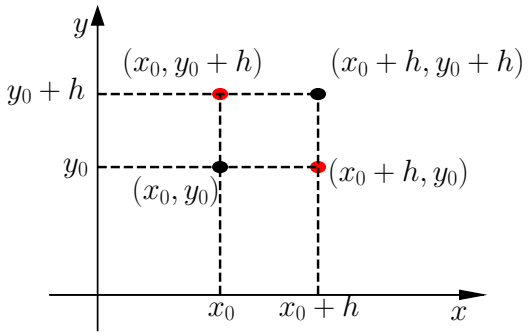
$$f_{xx}(x,y) = 2y^2, \qquad f_{yy}(x,y) = -x\sin(y) + 2x^2,$$
$$f_{xy}(x,y) = \cos(y) + 4xy, \qquad f_{yx}(x,y) = \cos(y) + 4xy.$$

Las derivadas de orden superior siguen la misma lógica. Así tenemos por ejemplo $f_{xxy}(x,y) = 4y$.

En el ejemplo observamos que las derivadas parciales cruzadas coinciden. Es decir, $f_{xy} = f_{yx}$, en todo punto. Veamos que este comportamiento es bastante general:

Teorema 6.23 (Schwarz)**.** Si f_{xy} y f_{yx} existen en un $B((x_0,y_0),\delta)$ y son continuas en (x_0,y_0), entonces $f_{xy}(x_0,y_0) = f_{yx}(x_0,y_0)$.

Demostración.



Cuando trabajamos con derivadas de primer orden, básicamente consideramos diferencias del estilo $f(a+h) - f(a)$. Ahora, al trabajar con derivadas de segundo orden, consideraremos el siguiente incremento que tiene información de cuatro puntos:

$$\varphi(h) = f(x_0+h, y_0+h) - f(x_0+h, y_0) - f(x_0, y_0+h) + f(x_0, y_0).$$

Ordenemos esta suma de dos formas distintas:

$$\varphi(h) = \Big[f(x_0+h, y_0+h) - f(x_0+h, y_0)\Big] - \Big[f(x_0, y_0+h) - f(x_0, y_0)\Big], \tag{6.2}$$
$$\varphi(h) = \Big[f(x_0+h, y_0+h) - f(x_0, y_0+h)\Big] - \Big[f(x_0+h, y_0) - f(x_0, y_0)\Big]. \tag{6.3}$$

Trabajemos primero con la expresión (6.2), y observemos que en cada uno de los dos paréntesis, tenemos la diferencia de f evaluada en dos puntos, donde además la segunda coordenada vale y_0+h e y_0. Hagamos uso entonces de una función auxiliar, que llamaremos ψ, definida de la siguiente manera: $\psi(x) = f(x, y_0+h) - f(x, y_0)$.

Entonces tenemos que

$$\begin{aligned}\varphi(h) &= \overbrace{\Big[f(x_0+h, y_0+h) - f(x_0+h, y_0)\Big]}^{\psi(x_0+h)} - \overbrace{\Big[f(x_0, y_0+h) - f(x_0, y_0)\Big]}^{\psi(x_0)} \\ &= \psi(x_0+h) - \psi(x_0).\end{aligned}$$

Podemos usar ahora el Teorema de Valor Medio de Lagrange para ψ. Resulta:

$$\varphi(h) = \psi(x_0+h) - \psi(x_0) = h\psi'(x_0+\theta h),$$

con $\theta \in (0,1)$. Ahora, como habíamos definido $\psi(x) = f(x, y_0+h) - f(x,y_0)$, su derivada es $\psi'(x) = f_x(x, y_0+h) - f_x(x,y_0)$, y por lo tanto tenemos:

$$\varphi(h) = h\big[f_x(x_0+\theta h, y_0+h) - f_x(x_0+\theta h, y_0)\big].$$

Y ahora observemos que tenemos la diferencia de la función f_x, evaluada en dos puntos con idéntica coordenada x. Podemos usar nuevamente una función auxiliar, $g(y) = f_x(x_0+\theta h, y)$. Entonces:

$$\varphi(h) = h\big[\overbrace{f_x(x_0+\theta h, y_0+h)}^{g(y_0+h)} - \overbrace{f_x(x_0+\theta h, y_0)}^{g(y_0)}\big] = h\big[g(y_0+h) - g(y_0)\big].$$

Usando nuevamente el Teorema de Valor Medio, ahora para g, resulta:

$$\varphi(h) = h\big[g(y_0+h) - g(y_0)\big] = h\big[hg'(y_0+\theta_2 h)\big],$$

con $\theta_2 \in (0,1)$. Nuevamente, como $g(y) = f_x(x_0+\theta h, y)$, su derivada es $g'(y) = f_{xy}(x_0+\theta h, y)$. Por lo tanto,

$$\varphi(h) = h^2 g'(y_0+\theta_2 h) = h^2 f_{xy}(x_0+\theta h, y_0+\theta_2 h).$$

Como f_{xy} es continua en (x_0,y_0), tenemos que cuando h tiende a cero:

$$\lim \frac{\varphi(h)}{h^2} = f_{xy}(x_0,y_0).$$

Partiendo de la expresión (6.3) y realizando el mismo procedimiento, se llega a que

$$\lim \frac{\varphi(h)}{h^2} = f_{yx}(x_0,y_0).$$

Luego, como el límite debe ser único, se concluye que $f_{xy}(x_0,y_0) = f_{yx}(x_0,y_0)$. □

Además de probar el resultado deseado, en el correr de esta demostración llegamos a un hecho interesante: encontramos una forma de aproximar numéricamente las derivadas cruzadas. Es decir, si en una computadora necesitamos el valor de una derivada de segundo orden f_{xy}, tomando un h pequeño y calculando $\frac{\varphi(h)}{h^2}$, obtenemos una aproximación de la derivada buscada. Esto, que es muy útil para aplicaciones en ingeniería, será estudiado en futuros cursos.

Volviendo al resultado en sí, una pregunta que surge naturalmente es si vale el recíproco de este teorema. Es decir, si f es una función tal que existen sus derivadas de segundo orden, y las derivadas cruzadas coinciden, ¿entonces necesariamente estas derivadas son continuas?. El siguiente ejemplo muestra que esto es falso.

Ejercicio 6.24

Consideremos la función: $f(x,y) = \begin{cases} xy^2 \sin\left(\frac{1}{y}\right) & si\ y \neq 0 \\ 0 & si\ y = 0. \end{cases}$

Comprobar que $f_{xy}(0,0) = f_{yx}(0,0) = 0$*, pero* f_{xy} *y* f_{yx} *no son continuas en* $(0,0)$.

Vale preguntarse también si las derivadas cruzadas, si existen, coinciden siempre. Es decir, si es realmente necesario pedir que estas derivadas sean continuas en el punto. El ejemplo que sigue muestra un función cuyas derivadas segundas f_{xy} y f_{yx} no coinciden en un punto.

Ejercicio 6.25

Consideremos la función: $f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & si\ (x,y) \neq (0,0) \\ 0 & si\ (x,y) = (0,0). \end{cases}$

Comprobar que $f_{xy}(0,0) = -1$ *y que* $f_{yx}(0,0) = 1$.

6.4. Funciones a $\mathbb{R}^m$

Hasta ahora hemos trabajado con funciones de $\mathbb{R}^n$ a $\mathbb{R}$, con especial énfasis en funciones definidas en $\mathbb{R}^2$. Veamos las generalizaciones correspondientes para funciones $f : \mathbb{R}^n \to \mathbb{R}^m$.

En estos casos, la imagen es un punto de $\mathbb{R}^m$, y por lo tanto tenemos m funciones (una para cada coordenada), y cada una de ellas depende de las n variables. Es decir,

$$f(x_1,x_2,\ldots,x_n) = \Big(f_1(x_1,x_2,\ldots,x_n), f_2(x_1,x_2,\ldots,x_n),\ldots,f_m(x_1,x_2,\ldots,x_n)\Big),$$

donde $f_i : \mathbb{R}^n \to \mathbb{R}$.

Ejemplo 6.26

Veamos un ejemplo de función $f : \mathbb{R}^3 \to \mathbb{R}^2$.

$$f(x,y,z) = \big(x + ye^z, \cos(x+y) - z\big).$$

Con la noción de continuidad no tenemos mayores problemas: como sabemos medir distancias tanto en el dominio como en el codominio, la definición dada en el capítulo anterior funciona sin variaciones.

La definición de diferenciabilidad también se puede generalizar fácilmente, pero de todas maneras la escribiremos para este caso:

Definición 6.27. Decimos que $f : \mathbb{R}^n \to \mathbb{R}^m$ es diferenciable en un punto $a \in \mathbb{R}^n$ sii existe una transformación lineal $df_a : \mathbb{R}^n \to \mathbb{R}^m$ tal que

$$f(a+h) = f(a) + df_a(h) + r(h), \quad \text{con } \lim_{h\to 0} \frac{r(h)}{\|h\|} = 0.$$

Observemos que aquí h es un incremento en $\mathbb{R}^n$, y la función resto, $r(h)$, va de $\mathbb{R}^n$ a $\mathbb{R}^m$, y por lo tanto el cero que aparece en la condición del resto, es el cero de $\mathbb{R}^m$.

Recordemos el primer resultado importante que vimos sobre diferenciabilidad, el 6.13. Allí, trabajando con funciones de $\mathbb{R}^2$ a $\mathbb{R}$, vimos que los elementos de la transformación lineal que aparece en la definición de diferenciabilidad, no son otros que las derivadas parciales en el punto.

En este caso, lo que obtenemos (con las mismas cuentas que en el Teorema 6.13), es que la matriz asociada a la transformación lineal df_a es la que tiene a las derivadas parciales de cada función, respecto a cada variable.

Es decir, es una matriz de m filas y n columnas (esto ya lo sabíamos, por el dominio y codominio de la transformación lineal), y en la entrada (i,j) de la matriz encontramos la derivada de la función f_i respecto de la variable x_j. Esta matriz lleva el nombre de *matriz Jacobiana*, y se denota como $J_f(a)$, es decir la matriz Jacobiana de la función f en el punto a. Tenemos entonces:

$$J_f(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}.$$

Dicho de otra manera, la matriz Jacobiana contiene a los gradientes de las funciones f_i en cada una de sus filas:

$$J_f(a) = \begin{pmatrix} \nabla f_1(a) \\ \hline \nabla f_2(a) \\ \hline \vdots \\ \hline \nabla f_m(a) \end{pmatrix}.$$

Ejemplo 6.28
Sea $f : \mathbb{R}^3 \to \mathbb{R}^2$ *definida como* $f(x,y,z) = \left(x^2y + e^z, \sin(x) + yz\right)$. *En este caso, tenemos* $f_1(x,y,z) = x^2y + e^z$, *y por otro lado* $f_2(x,y,z) = \sin(x) + yz$.

Entonces la matriz Jacobiana en un punto genérico (x,y,z) *es:*

$$J_f(x,y,z) = \begin{pmatrix} 2xy & x^2 & e^z \\ \cos(x) & z & y \end{pmatrix}.$$

Veamos ahora entonces la versión general de la Regla de la Cadena.

Teorema 6.29 (Regla de la Cadena III)**.** Sean $g : \mathbb{R}^k \to \mathbb{R}^n$, diferenciable en $a \in \mathbb{R}^k$, y $f : \mathbb{R}^n \to \mathbb{R}^m$, diferenciable en $g(a)$. Entonces $f \circ g : \mathbb{R}^k \to \mathbb{R}^m$ es diferenciable en a, y además su diferencial es $d(f \circ g)_a = df_{g(a)} \circ dg_a$.

Observar que el diferencial de la composición es la composición de los diferenciales (evaluados en los puntos correctos), y por lo tanto, la matriz Jacobiana de $f \circ g$ es el producto de las matrices Jacobianas de f y g (recordar que el producto matricial no es conmutativo, por lo tanto hay que tener cuidado con el orden): $J_{f\circ g}(a) = J_f(g(a)) \cdot J_g(a)$.

Demostremos entonces esta versión de la Regla de la Cadena.

Demostración. Para demostrar la diferenciabilidad de la $f \circ g$ en el punto a, veamos que verifica la definición. Entonces veamos qué sucede con $(f \circ g)(a+h)$.

$$(f \circ g)(a+h) = f(g(a+h)) = f\Big(g(a) + dg_a(h) + r_g(h)\Big)$$

donde usamos la diferenciabilidad de g en el punto a. Ahora, observemos que tenemos evaluada la función f en el punto $g(a)$ (donde la función f es diferenciable), más un incremento, que llamaremos $v = dg_a(h) + r_g(h)$. Es decir:

$$f(g(a+h)) = f\Big(g(a) + \underbrace{dg_a(h) + r_g(h)}_{v}\Big) = f\Big(g(a) + v\Big).$$

Pero como f es diferenciable en $g(a)$, tenemos que:

$$f(g(a+h)) = f\Big(g(a) + v\Big) = f(g(a)) + df_{g(a)}(v) + r_f(v).$$

Ahora, recordando que v era el incremento $v = dg_a(h) + r_g(h)$, recuperamos:

$$f(g(a+h)) = f\Big(g(a)\Big) + df_{g(a)}\Big(dg_a(h) + r_g(h)\Big) + r_f\Big(dg_a(h) + r_g(h)\Big).$$

Como el diferencial $df_{g(a)}$ es una transformación lineal (por definición), tenemos que $df_{g(a)}\Big(dg_a(h) + r_g(h)\Big) = df_{g(a)}\Big(dg_a(h)\Big) + df_{g(a)}\Big(r_g(h)\Big)$. Llegamos entonces a una expresión muy auspiciosa para lo que queremos demostrar, ya que tenemos:

$$\begin{aligned} f(g(a+h)) &= f\Big(g(a)\Big) + df_{g(a)}\Big(dg_a(h)\Big) + df_{g(a)}\Big(r_g(h)\Big) + r_f\Big(dg_a(h) + r_g(h)\Big) \\ &= f\Big(g(a)\Big) + \Big(df_{g(a)} \circ dg_a\Big)(h) + \underbrace{df_{g(a)}\Big(r_g(h)\Big) + r_f\Big(dg_a(h) + r_g(h)\Big)}_{R(h)}. \end{aligned}$$

Es decir, la función $f \circ g$ evaluada en $a+h$ la podemos escribir como la función en el punto a, más una transformación lineal (que es la composición de los diferenciales de f y g) evaluada en el incremento h, más un resto $R(h)$. Solo resta probar que $R(h)$ tiende a cero más rápido que h. Es decir, que:

$$\lim_{h\to 0} \frac{R(h)}{\|h\|} = 0.$$

El resto $R(h)$ es todo lo que "sobra" de la aproximación lineal a $f \circ g$. Es decir:

$$R(h) = df_{g(a)}\Big(r_g(h)\Big) + r_f\Big(dg_a(h) + r_g(h)\Big).$$

Debemos probar entonces que

$$\lim_{h\to 0} \frac{df_{g(a)}\Big(r_g(h)\Big)}{\|h\|} + \frac{r_f\Big(dg_a(h) + r_g(h)\Big)}{\|h\|} = 0.$$

Para el primer término, basta observar que como el diferencial es una transformación lineal, podemos incluir la norma de h en el argumento. Es decir,

$$\frac{df_{g(a)}\Big(r_g(h)\Big)}{\|h\|} = df_{g(a)}\left(\frac{r_g(h)}{\|h\|}\right).$$

Luego, como $\frac{r_g(h)}{\|h\|} \to 0$ (porque esa es la propiedad que sabemos del resto r_g, a partir de la diferenciabilidad de g), y las transformaciones lineales son continuas, concluimos que

$$\lim_{h\to 0} df_{g(a)}\left(\frac{r_g(h)}{\|h\|}\right) = 0.$$

Falta ver que el otro sumando también tiende a cero. Para esto, volvemos a llamar v al argumento de r_f, es decir, $v = dg_a(h) + r_g(h)$. Luego tenemos:

$$\lim_{h\to 0} \frac{r_f\Big(dg_a(h) + r_g(h)\Big)}{\|h\|} = \lim_{h\to 0} \frac{r_f\Big(v\Big)}{\|h\|} = \lim_{h\to 0} \frac{r_f\Big(v\Big)}{\|v\|} \frac{\|v\|}{\|h\|},$$

donde multiplicamos y dividimos por $\|v\|$, para poder utilizar la propiedad del resto r_f. Queda como ejercicio demostrar que cuando $h \to 0$, tenemos que $v \to 0$, y que $\frac{\|v\|}{\|h\|}$ está acotado, para concluir que

$$\lim_{h\to 0} \underbrace{\frac{r_f\Big(v\Big)}{\|v\|}}_{\to 0} \underbrace{\frac{\|v\|}{\|h\|}}_{acotado} = 0.$$

□

Ejemplo 6.30
Consideremos por un lado la función $f : \mathbb{R}^3 \to \mathbb{R}^2$ definida como $f(x,y,z) = \big(x^2y + e^z, \sin(x) + yz\big)$, de la cual calculamos su matriz Jacobiana en el ejemplo 6.28.

Por otro lado, sea $g : \mathbb{R}^2 \to \mathbb{R}^3$ definida como $g(u,v) = \big(uv, e^u, \cos(v)\big)$, y llamemos h a la función que resulta de componerlas. Es decir, $h : \mathbb{R}^2 \to \mathbb{R}^2$ es $h = f \circ g$, o lo que es lo mismo, $h(u,v) = f\big(g(u,v)\big)$.

Calculemos la matriz Jacobiana de h en el punto $(u,v) = (1,\pi)$.

En este caso, como conocemos las funciones f y g en su totalidad, podemos calcular explícitamente la composición y luego derivar. Sigamos primero este camino entonces.

Tenemos que $h(u,v) = \left(u^2v^2e^u + e^{\cos(v)}, \sin(uv) + e^u\cos(v)\right)$, y por lo tanto su matriz Jacobiana en un punto genérico (u,v) es:[3]

$$J_h(u,v) = \begin{pmatrix} e^u v^2 u(2+u) & 2ve^u u^2 - \sin(v)e^{\cos(v)} \\ v\cos(uv) + e^u\cos(v) & u\cos(uv) - e^u\sin(v) \end{pmatrix}.$$

Evaluando ahora en el punto $(1,\pi)$, concluimos que

$$J_h(1,\pi) = \begin{pmatrix} 3e\pi^2 & 2\pi e \\ -\pi - e & -1 \end{pmatrix}.$$

Calculemos nuevamente esta matriz, pero ahora utilizando la regla de la cadena. Esto es:

$$J_{f\circ g}(1,\pi) = J_f(g(1,\pi)) \cdot J_g(1,\pi).$$

Observemos que $g(1,\pi) = (\pi, e, -1)$. Luego, como ya tenemos calculada la matriz Jacobiana de f, del ejemplo 6.28, tenemos que

$$J_f(\pi, e, -1) = \begin{pmatrix} 2\pi e & \pi^2 & e^{-1} \\ -1 & -1 & e \end{pmatrix}.$$

Por otro lado, la matriz Jacobiana de g es:

$$J_g(u,v) = \begin{pmatrix} v & u \\ e^u & 0 \\ 0 & -\sin(v) \end{pmatrix}.$$

Evaluando en $(1,\pi)$, tenemos

$$J_g(1,\pi) = \begin{pmatrix} \pi & 1 \\ e & 0 \\ 0 & 0 \end{pmatrix}.$$

Y ahora simplemente multiplicando las matrices, obtenemos

$$J_{f\circ g}(1,\pi) = \begin{pmatrix} 2\pi e & \pi^2 & e^{-1} \\ -1 & -1 & e \end{pmatrix} \cdot \begin{pmatrix} \pi & 1 \\ e & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 3e\pi^2 & 2\pi e \\ -\pi - e & -1 \end{pmatrix},$$

que coincide con el resultado calculado anteriormente.

[3] Ejercicio: comprobar los cálculos.

6.5. Desarrollo de Taylor

En esta sección veremos la versión para varias variables del Teorema de Taylor, que ya fue estudiado para el caso de funciones a valores reales.

Por un lado, es un resultado interesante en sí mismo, ya que nos asegura que, bajo ciertas condiciones, podemos aproximar (localmente) cualquier función a través de polinomios.

Por otro lado, es una herramienta sumamente útil en cualquier aplicación, en ingeniería en particular, ya que muchas veces somos capaces de hacer cálculos, resolver problemas, solamente en casos simples (por ejemplo para funciones lineales, o para ecuaciones de orden dos, etc). Entonces, en esos casos, se puede aproximar una función por su polinomio de Taylor. De esta manera somos capaces de obtener una solución aproximada, pero además somos capaces de *controlar el error*.

Comencemos con los resultados conocidos en una variable.

Repaso en $\mathbb{R}$

Si $f : \mathbb{R} \to \mathbb{R}$ es una función C^{k+1}, entonces

$$f(x_0+h) = f(x_0) + f'(x_0)h + \frac{f''(x_0)}{2}h^2 + \frac{f'''(x_0)}{3!}h^3 + \cdots + \frac{f^{(k)}(x_0)}{k!}h^k + r(h),$$

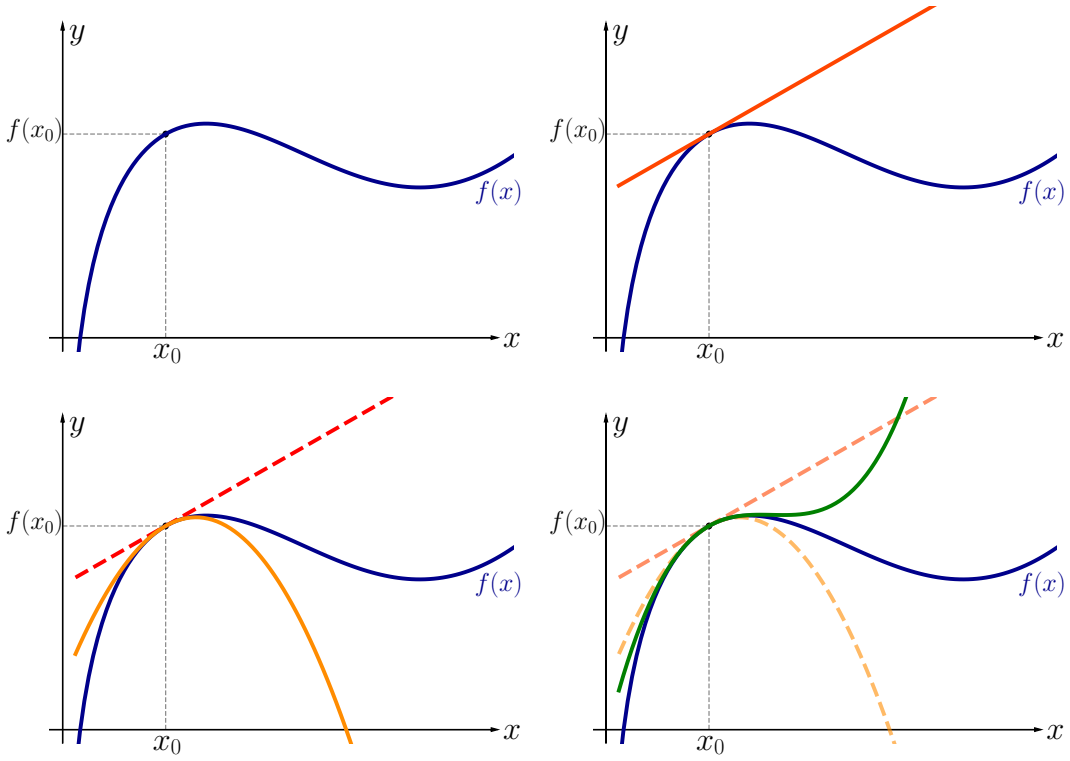
con r una función que cumple $\lim_{h\to 0} \frac{r(h)}{h^k} = 0$.

Es decir, localmente (cuando h tiende a cero) la función puede ser aproximada por un polinomio.

Observar que con los primeros dos términos de la aproximación (o mirando el desarrollo de Taylor con $k = 1$) tenemos la definición de diferenciabilidad.

La propiedad del resto nos dice que el error de la aproximación tiende a cero más rápido que h^k. Esto, que muchas veces es olvidado en un enunciado rápido del resultado, es de lo más importante del teorema (ver discusión sobre le resto en la definición (6.1)).

La interpretación geométrica en una variable se puede ver en la figura: con orden 1 obtenemos la recta que mejor aproxima la función en un punto, a medida que vamos aumentando el orden, obtenemos mejores aproximaciones con parábolas, polinomios de tercer orden, etc.



En varias variables, ya lo tenemos hecho para orden uno: es la definición de diferenciabilidad (6.11). Si pensamos en funciones de $\mathbb{R}^2$ a $\mathbb{R}$, esto equivale a aproximar la superficie del gráfico de la función por un plano, como ya vimos. Cuando utilizamos polinomios de órdenes mayores, estaremos aproximando el gráfico de f por otras superficies (como un paraboloide por ejemplo). Debemos ver cuál es el equivalente de los términos de mayor orden cuando trabajamos con varias variables.

Diferenciales de orden superior

Veremos cómo son los diferenciales de orden superior, y el desarrollo de Taylor, para funciones de dos variables.

Recordemos que si f es diferenciable en a, el diferencial es la transformación lineal que mejor aproxima localmente: $df_a(\Delta x, \Delta y) = f_x(a)\Delta x + f_y(a)\Delta y$.

Consideremos una función $f \in C^2$, definimos el diferencial segundo en a como la función $d^2 f_a : \mathbb{R}^2 \to \mathbb{R}$ definida por[4]:

$$d^2 f_a(\Delta x, \Delta y) = f_{xx}(a)\Delta x^2 + 2f_{xy}(a)\Delta x \Delta y + f_{yy}(a)\Delta y^2.$$

Como pedimos que la función sea de clase C^2, las derivadas cruzadas son iguales, y por eso aparece dos veces el término $f_{xy}(a)\Delta x\Delta y$ (ver la definición general en la nota al pie).

[4]En general, para funciones de $\mathbb{R}^n$ a $\mathbb{R}$, se define el diferencial segundo como $d^2 f_a(v) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) v_i v_j$.

Observemos que el diferencial segundo es un polinomio homogéneo de segundo orden en las variables $\Delta x, \Delta y$. Esto es: en cada sumando, la suma de los exponentes de Δx y Δy es dos.

Además, podemos escribir esta expresión de forma matricial, que por un lado resulta más compacta, y por otro lado nos permitirá utilizar algunas herramientas matriciales. En efecto, definamos la *matriz Hessiana*, que la denotamos H, así:

$$H = \begin{pmatrix} f_{xx}(a) & f_{xy}(a) \\ f_{yx}(a) & f_{yy}(a) \end{pmatrix}.$$

Entonces, si notamos al vector de incrementos $v = (\Delta x, \Delta y)$, resulta $df_a^2(v) = vHv^T$. Ejercicio: corroborar esta igualdad. Observemos que si las derivadas cruzadas son iguales, esta matriz resulta simétrica.

De manera similar se definen los diferenciales de orden superior. Por ejemplo, si todas las derivadas conmutan, tenemos:

$$d^3 f_a(\Delta x, \Delta y) = f_{xxx}(a)\Delta x^3 + 3f_{xxy}(a)\Delta x^2 \Delta y + 3f_{xyy}(a)\Delta x \Delta y^2 + f_{yyy}(a)\Delta y^3.$$

Aparecen todas las derivadas de orden tres, y los coeficientes, como antes, responden a la cantidad de formas que tenemos de "ordenar" las coordenadas respecto a las que derivamos. Por ejemplo, si f es de clase C^3, entonces las derivadas f_{xxy}, f_{xyx}, y f_{yxx} son iguales, y por eso aparece un 3 como coeficiente, que son las tres maneras de ordenar xxy.

Además...

En general, si p es un natural, y tenemos una función $f : \mathbb{R}^n \to \mathbb{R}$ con todas las derivadas de orden p, el diferencial de orden p se define como:

$$d^p f_a(v) = \sum_{i_1, i_2, \ldots, i_p = 1}^{n} \frac{\partial^p f}{\partial x_{i_1} \ldots \partial x_{i_p}}(a) v_{i_1} v_{i_2} \ldots v_{i_p}$$

donde $v = (v_1, v_2, \ldots, v_n)$ es el vector donde evaluamos df_a^p.
En el caso de funciones de dos variables de clase C^p, esto resulta:

$$d^p f_a(\Delta x, \Delta y) = \sum_{i=0}^{p} \binom{p}{i} \frac{\partial^p f}{\partial x^{p-i} \partial y^i}(a) \Delta x^{p-i} \Delta y^i$$
$$= \frac{\partial^p f}{\partial x^p}(a)\Delta x^p + \binom{p}{1} \frac{\partial^p f}{\partial x^{p-1} \partial y}(a)\Delta x^{p-1}\Delta y + \cdots + \binom{p}{i} \frac{\partial^p f}{\partial x^{p-i} \partial y^i}(a)\Delta x^{p-i}\Delta y^i + \cdots + \frac{\partial^p f}{\partial y^p}(a)\Delta y^p$$

donde $\binom{p}{i}$ es el coeficiente binomial, que cuenta la cantidad de formas que hay de elegir las i ubicaciones de (digamos) la y en los p lugares. Por ejemplo, como ya vimos, hay tres formas de ordenar dos x y una y: xxy, xyx, yxx. Este coeficiente $\binom{3}{2} = 3$, es el que aparece en el diferencial segundo. En el esquema de demostración del Teorema de Taylor podemos entender por qué aparecen estos números combinatorios.

Estamos en condiciones entonces de enunciar el Teorema de Taylor:

Teorema 6.31 (Taylor)**.** Sea $f:\mathbb{R}^m\to\mathbb{R}$ una función de clase C^{k+1} en un entorno de $a\in\mathbb{R}^m$. Entonces,

$$f(a+h)=f(a)+df_a(h)+\frac{1}{2}d^2f_a(h)+\frac{1}{3!}d^3f_a(h)+\cdots+\frac{1}{k!}d^kf_a(h)+r_k(h),$$

donde el resto r_k es una función que cumple $\lim\limits_{h\to(0,0,\ldots,0)}\dfrac{r_k(h)}{\|h\|^k}=0$.

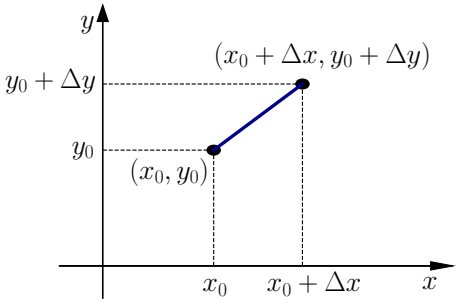
Por ejemplo, en el caso de una función de dos variables, el desarrollo de Taylor de orden dos resulta:

$$f(x_0+\Delta x,y_0+\Delta y)=$$
$$f(x_0,y_0)+f_x(x_0,y_0)\Delta x+f_y(x_0,y_0)\Delta y+\frac{1}{2}\left[f_{xx}(x_0,y_0)\Delta x^2+2f_{xy}(x_0,y_0)\Delta x\Delta y+f_{yy}(x_0,y_0)\Delta y^2\right]+r(\Delta x,\Delta y)$$

Veamos un esquema de la demostración de este Teorema, para el caso de dos variables.

Demostración.

Como queremos ver cómo se comporta la función al pasar del punto (x_0,y_0) al punto $(x_0+\Delta x,y_0+\Delta y)$, consideremos la función auxiliar $g:[0,1]\to\mathbb{R}$, definida como $g(t)=f(x_0+t\Delta x,y_0+t\Delta y)$. Es decir, nos movemos desde el punto original (x_0,y_0) (en $t=0$) al punto $(x_0+\Delta x,y_0+\Delta y)$ (en $t=1$) siguiendo el segmento que los une.



Como g es una función de una variable, que hereda toda la diferenciabilidad de f, podemos aplicar el Teorema de Taylor que ya conocemos para funciones de una variable. Para esto, debemos calcular las derivadas de g, usando la regla de la cadena (es un buen ejercicio intentar hallar esta derivada antes de ver el resultado):

$$g'(t)=f_x(x_0+t\Delta x,y_0+t\Delta y)\Delta x+f_y(x_0+t\Delta x,y_0+t\Delta y)\Delta y.$$

Como vamos a calcular el desarrollo de Taylor de g en $t=0$ (que corresponde a $f(x_0,y_0)$), evaluamos:

$$g'(0)=f_x(x_0,y_0)\Delta x+f_y(x_0,y_0)\Delta y,$$

que es exactamente el diferencial de f en el incremento $(\Delta x,\Delta y)$.

Luego, para calcular la derivada segunda, derivamos la expresión obtenida para $g'(t)$, respecto de la variable t por supuesto. Observar que t aparece cuatro veces en dicha expresión:

$$g'(t)=f_x(x_0+t\Delta x,y_0+t\Delta y)\Delta x+f_y(x_0+t\Delta x,y_0+t\Delta y)\Delta y.$$

Veamos la derivada del primer sumando:

$$\frac{\partial f_x(x_0+t\Delta x, y_0+t\Delta y)}{\partial t} = f_{xx}(x_0+t\Delta x, y_0+t\Delta y)\Delta x + f_{xy}(x_0+t\Delta x, y_0+t\Delta y)\Delta y.$$

De manera similar, tenemos:

$$\frac{\partial f_y(x_0+t\Delta x, y_0+t\Delta y)}{\partial t} = f_{yx}(x_0+t\Delta x, y_0+t\Delta y)\Delta x + f_{yy}(x_0+t\Delta x, y_0+t\Delta y)\Delta y.$$

Observar que en la derivada del primer sumando aparece f_{xy}, y en la derivada del segundo sumando aparece f_{yx}. Como la función es C^2, estas derivadas son iguales, y por lo tanto evaluando en $t=0$ y sumando los términos tenemos:

$$g''(0) = f_{xx}(x_0,y_0)\Delta x^2 + 2f_{xy}(x_0,y_0)\Delta x\Delta y + f_{yy}(x_0,y_0)\Delta y^2,$$

que es exactamente el diferencial segundo de f:

$$g''(0) = d^2 f_{(x_0,y_0)}(\Delta x, \Delta y).$$

De igual manera se calculan las derivadas siguientes, y van apareciendo los números combinatorios, de la misma forma que apareció el 2 al combinar f_{xy} y f_{yx}. Por ejemplo, al derivar el término de $g''(t)$ que contiene $f_{xx}(x_0+t\Delta x, y_0+t\Delta y)$, obtendremos dos términos: uno con f_{xxx} y otro con f_{xxy}. La derivada respecto a x tres veces es la única que aparecerá en el cálculo de $g'''(t)$, pero la derivada f_{xxy} aparecerá dos veces más (¿dónde?).

Observando la relación entre las derivadas de g y los diferenciales de f, y utilizando el Teorema de Taylor en una variable, se concluye la demostración. □

Veamos un ejemplo.

Ejemplo 6.32
Calculemos el desarrollo de Taylor de segundo orden de $f(x,y) = \log(1+2x+y)$ *alrededor del punto* $a = (1,2)$. *Primero calculamos* $f(1,2) = \log(5)$.

Veamos ahora las derivadas:

$$f_x(x,y) = \frac{2}{1+2x+y} \Longrightarrow f_x(1,2) = \frac{2}{5}$$

$$f_y(x,y) = \frac{1}{1+2x+y} \Longrightarrow f_y(1,2) = \frac{1}{5}$$

$$f_{xx}(x,y) = \frac{-4}{(1+2x+y)^2} \Longrightarrow f_{xx}(1,2) = \frac{-4}{25}$$

$$f_{xy}(x,y) = \frac{-2}{(1+2x+y)^2} \Longrightarrow f_{xx}(1,2) = \frac{-2}{25}$$

$$f_{yy}(x,y) = \frac{-1}{(1+2x+y)^2} \Longrightarrow f_{xx}(1,2) = \frac{-1}{25}$$

Por lo tanto el desarrollo de Taylor resulta:

$$f(1+\Delta x, 2+\Delta y) = \log(5) + \frac{2}{5}\Delta x + \frac{1}{5}\Delta y + \frac{1}{2}\left[-\frac{4}{25}\Delta x^2 - \frac{4}{25}\Delta x\Delta y - \frac{1}{25}\Delta y^2\right] + r(\Delta x, \Delta y),$$

con $\lim_{(\Delta x,\Delta y)\to(0,0)} \dfrac{r(\Delta x,\Delta y)}{\|(\Delta x,\Delta y)\|^2} = 0.$

Equivalentemente también podemos escribir:

$$f(x,y) = \log(5) + \frac{2}{5}(x-1) + \frac{1}{5}(y-2) - \frac{1}{50}\left[4(x-1)^2 + 4(x-1)(y-2) + (y-2)^2\right] + r(x-1, y-2),$$

con la misma propiedad del resto.

6.6. Extremos relativos

Un problema muy importante donde se utiliza el desarrollo de Taylor, es el de hallar máximos y mínimos de una función.

Dada una función continua f definida en un conjunto compacto, el Teorema de Weierstrass nos asegura que tendrá máximo y mínimo absolutos. Los puntos donde se alcanzan pueden estar en el interior del conjunto o en el borde, y para hallarlos se utilizan herramientas distintas dependiendo de si buscamos puntos interiores o en la frontera. En esta sección veremos un primer acercamiento al comportamiento en puntos interiores, y más adelante se estudiará el problema en su totalidad.

Definición 6.33. Consideremos una función $f : D \to \mathbb{R}$, con $D \subset \mathbb{R}^n$, y a un punto interior a D. Decimos que f tiene un máximo relativo en a sii $\exists \delta > 0$ tal que $f(a) \geqslant f(x)$ para todo punto $x \in B(a,\delta) \cap D$.
De manera análoga se define mínimo relativo.

Se denomina *extremos relativos* a los máximos y mínimos relativos.

Proposición 6.34. Si $f : D \subset \mathbb{R}^n \to \mathbb{R}$ es diferenciable y tiene un extremo relativo en $a \in \mathring{D}$, entonces $\nabla f(a) = (0,0,\ldots,0)$.

Demostración. Basta restringir a cada dirección canónica, y aplicar el resultado conocido en una variable a la derivada parcial correspondiente. □

A los puntos donde se anula el gradiente se los llama *puntos críticos*.

Observación 6.35

- *Recordar que en una variable, esta condición no es suficiente. Es decir, puede haber puntos donde se anule la derivada, pero donde la función no tenga un extremo relativo. El ejemplo típico es $f(x) = x^3$ en el origen.*
- *En varias variables esto sigue siendo cierto. Se pueden construir ejemplos similares utilizando potencias impares (como por ejemplo $f(x,y) = x^3 + y^3$), o se puede hacer uso de las diferentes direcciones, como por ejemplo $f(x,y) = x^2 - y^2$. Ejercicio: comprobar estos ejemplos.*

Si tenemos una función diferenciable, y nos proponemos encontrar los extremos relativos en puntos interiores, entonces debemos buscar todos los puntos críticos, y luego clasificarlos: esto es, decidir si cada uno de los puntos críticos es un máximo relativo, un mínimo relativo, o ninguno de los dos.

Como para cada punto crítico buscamos entender el comportamiento local, lo que haremos es estudiar su desarrollo de Taylor.

Específicamente, supongamos que tenemos un punto crítico a, interior al dominio. Para saber si es un extremo relativo, debemos estudiar cómo es el valor funcional de los puntos cercanos a a, respecto a $f(a)$. Esto es, miraremos el signo de $f(a+h) - f(a)$. Escribimos entonces el desarrollo de Taylor de f alrededor de a:

$$f(a+h) = f(a) + df_a(h) + \frac{1}{2}d^2 f_a(h) + r(h).$$

Pero como a es un punto crítico, el diferencial primero se anula. Por lo tanto, tenemos que:

$$f(a+h) - f(a) = \frac{1}{2}d^2 f_a(h) + r(h),$$

y en definitiva el signo de $f(a+h) - f(a)$ es el signo del diferencial segundo[5]. Geométricamente, lo que estamos viendo es: dado un punto crítico, donde el plano que mejor aproxima a la función es horizontal, miramos si la aproximación de segundo orden es un paraboloide ("hacia arriba" o "hacia abajo"), o es una silla de montar, como la de la Figura 5.4. En este último caso, tendremos lo que se denomina un *punto silla*, y sino será un extremo relativo. En realidad, puede que la información del diferencial segundo no sea suficiente para clasificar el punto crítico, como veremos en breve.

Recordemos que el diferencial segundo es un polinomio homogéneo de segundo grado, que además puede escribirse en forma matricial:

$$d^2 f_a(\Delta x, \Delta y) = f_{xx}(a)\Delta x^2 + 2f_{xy}(a)\Delta x\Delta y + f_{yy}(a)\Delta y^2 = (\Delta x, \Delta y)H(\Delta x, \Delta y)^T.$$

[5]Hay que ver que efectivamente el resto no influye localmente en el signo, y bajo qué condiciones. Por ejemplo, si todas las derivadas segundas son nulas, claramente un estudio de signo basado en la aproximación de segundo orden fallará.

Supongamos primero que $f_{xy}(a) = 0$, lo que corresponde a que la matriz Hessiana sea diagonal. Entonces tenemos que

$$d^2 f_a(\Delta x, \Delta y) = f_{xx}(a)\Delta x^2 + f_{yy}(a)\Delta y^2.$$

Como los términos Δx^2 y Δy^2 son siempre no negativos, podemos estudiar el signo de $d^2 f_a(\Delta x, \Delta y)$ a partir del signo de las derivadas f_{xx} y fyy. Por ejemplo, si ambas derivadas son estrictamente positivas, entonces $d^2 f_a(\Delta x, \Delta y) > 0$ para todo $(\Delta x, \Delta y) \neq (0,0)$. Esto corresponde a un mínimo relativo en a, ya que localmente, $f(a+h) - f(a)$ es positivo. Análogamente, si ambas derivadas son estrictamente negativas, tenemos un máximo relativo en a. Si tenemos una derivada positiva y una negativa, entonces en una dirección canónica la función crece, y en la otra dirección la función decrece, por lo que no puede haber un extremo relativo en a, sino que es un punto silla. Si, además de f_{xy}, otra derivada es cero, entonces no podemos concluir nada a partir del diferencial segundo[6].

En esta discusión asumimos una matriz Hessiana diagonal, y en realidad lo que observamos es el signo de las entradas de la diagonal, que no son otras que los valores propios de H. Con argumentos básicos de álgebra lineal (básicamente haciendo el cambio de coordenadas correspondiente a los vectores propios), es fácil ver que esta discusión es más general, y son los valores propios de H los que determinan el comportamiento del diferencial segundo. Específicamente, si los valores propios son λ_1 y λ_2:

- Si $\lambda_1 > 0$ y $\lambda_2 > 0$, entonces f tiene un mínimo relativo en a.
- Si $\lambda_1 < 0$ y $\lambda_2 < 0$, entonces f tiene un máximo relativo en a.
- Si $\lambda_1 > 0$ y $\lambda_2 < 0$, entonces a es un un punto silla.
- Si $\lambda_1 = 0$ o $\lambda_2 = 0$, entonces el diferencial segundo no clasifica.

Como en el caso de dimensión dos, el determinante y la traza determinan los signos de los valores propios, se puede presentar una clasificación también en estos términos:

- Si $\det(H) > 0$ y $tr(H) > 0$, entonces f tiene un mínimo relativo en a.
- Si $\det(H) > 0$ y $tr(H) < 0$, entonces f tiene un máximo relativo en a.
- Si $\det(H) < 0$, entonces a es un un punto silla.
- Si $\det(H) = 0$, entonces el diferencial segundo no clasifica.

[6] Es como si, en una variable, tuviéramos una función con derivadas primera y segunda nulas. Podría ser localmente como x^3, y no presentar extremo relativo, o si la derivada tercera también se anula, podría ser localmente como x^4 y sí presentar extremo relativo.

Esta clasificación por medio de los valores propios es más general, y podemos utilizarla también para el caso de dimensión n (si todos los valores propios tienen el mismo signo, si existen valores propios de signos opuestos, o si cero es un valor propio). No demostraremos aquí estos resultados. El estudio más detallado sobre clasificación de formas cuadráticas se puede ver en el curso de álgebra lineal (ver [3]).

Comencemos con ejemplos que corroboran lo que ya sabemos.

Ejemplos 6.36

- *Consideremos $f(x,y) = x^2 + y^2$. Entonces el gradiente es $\nabla f(x,y) = (2x, 2y)$, y por lo tanto el único punto crítico es el $(0,0)$. La matriz Hessiana en el origen (y en realidad en todo punto) es:*

$$H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

  por lo que f presenta un mínimo relativo en el origen, como ya sabíamos.

- *Consideremos ahora $f(x,y) = x^2 - y^2$. Entonces el gradiente es $\nabla f(x,y) = (2x, -2y)$, y también el único punto crítico es el $(0,0)$. La matriz Hessiana es:*

$$H = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix},$$

  por lo que el origen es un punto silla de f, y en realidad el nombre punto silla proviene de este ejemplo, que ya habíamos denominado silla de montar, en el ejercicio 5.2.

Observación 6.37
Consideremos ahora la función $f(x,y) = x^2 + y^3$, que tiene un punto crítico en el $(0,0)$. La matriz Hessiana en el origen es:

$$H = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

por lo tanto no podemos clasificar el punto crítico con el diferencial segundo. Observar que en este caso el $(0,0)$ es un punto silla (¿por qué?).

Ahora bien, la función $f(x,y) = x^2 + y^4$ también tiene el origen como único punto crítico, y tiene la misma matriz Hessiana. Sin embargo, en este caso es claro que la función presenta un mínimo relativo en el origen.

Es decir, la misma matriz Hessiana, cuando presenta valores propios nulos, puede corresponder a cualquier tipo de puntos críticos: extremos relativos o puntos silla.

Ejemplo 6.38
Consideremos la función $f:\mathbb{R}^2\to\mathbb{R}$ *definida por* $f(x,y)=x^3+y^3-3xy$. *El gradiente de* f *es* $\nabla f(x,y)=(3x^2-3y,3y^2-3x)$. *Por lo tanto, los puntos críticos son los que cumplen simultáneamente* $y=x^2$ *y* $x=y^2$. *Es fácil ver que los únicos puntos críticos son* $(0,0)$ *y* $(1,1)$. *Para clasificar estos puntos, calculemos la matriz Hessiana:*

$$H(x,y)=\begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}.$$

Entonces, para el punto $(0,0)$, *la matriz Hessiana resulta* $H(0,0)=\begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}$, *que tiene valores propios* 3 *y* -3. *Por lo tanto, la función tiene un punto silla en el origen.*

Por otro lado, para el punto $(1,1)$, *la matriz Hessiana resulta* $H(1,1)=\begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}$, *que tiene valores propios* 9 *y* 3. *Por lo tanto, la función tiene un mínimo relativo en el* $(1,1)$. *También bastaba observar que el determinante y la traza son positivos.*

Capítulo 7

Integrales múltiples

Repaso en una variable

Para empezar, recordemos muy brevemente la definición de integral en una variable, y veamos qué elementos podemos reutilizar directamente, y qué elementos requieren una nueva definición, en el contexto de más variables.

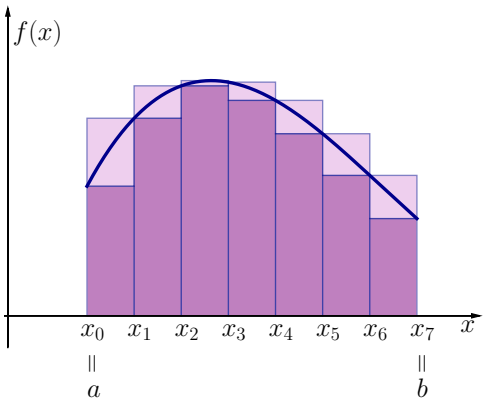
Cuando definimos la integral de una función acotada, en un intervalo $[a,b]$, nos basamos en la idea de aproximar el área bajo la curva mediante sumas de áreas de rectángulos. Consideramos una partición del intervalo $[a,b]$, que definimos como $\mathscr{P} = \{x_0, x_1, x_2, \ldots, x_n\}$, con $x_i < x_{i+1}$, donde los extremos son $x_0 = a$ y $x_n = b$, y a cada intervalo de la partición lo denominamos $I_i = [x_i, x_{i+1}]$.

Luego, definimos las sumas inferiores y superiores respecto de esa partición:

$$s(f,\mathscr{P}) = \sum_i l(I_i) \inf_{x\in I_i} f(x),$$

$$S(f,\mathscr{P}) = \sum_i l(I_i) \sup_{x\in I_i} f(x),$$

donde $l(I_i)$ es la longitud del intervalo I_i, es decir, $l(I_i) = x_{i+1} - x_i$.

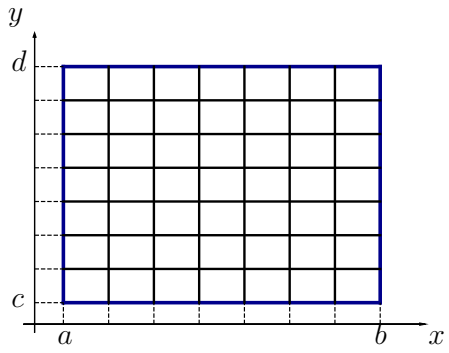


Cada término de la forma $l(I_i) \inf_{x\in I_i} f(x)$ corresponde al área del rectángulo de base $[x_i, x_{i+1}]$, que queda por debajo del gráfico de f en ese intervalo. Así, la suma inferior es la suma de estas áreas, y resulta una estimación *por defecto* de la integral de f.

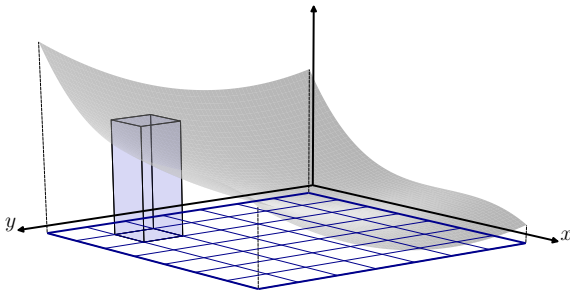
Lo mismo sucede con las sumas superiores. A medida que vamos refinando la partición,

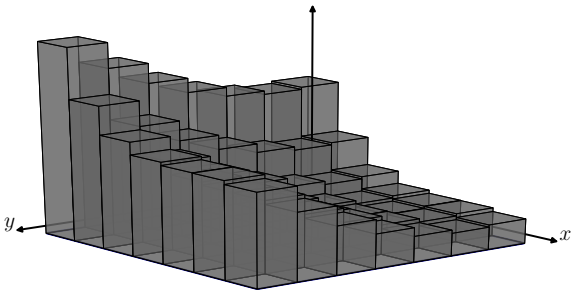
las sumas inferiores y superiores se van acercando entre sí, dando una mejor estimación del área bajo la curva. Cuando el supremo de las sumas inferiores (haciendo variar $\mathscr{P}$ en todas las particiones posibles) es igual al ínfimo de las sumas superiores, decimos que f es integrable en $[a,b]$, y que ese es el valor de la integral.

Intentaremos generalizar esta idea para una función de dos variables, y supongamos que está definida en un dominio rectangular D, para facilitar la descripción.



Una idea bastante natural es, en lugar de particionar el intervalo $[a,b]$ en pequeños intervalos, dividimos el rectángulo D en pequeños rectángulos. Para cada uno de estos pequeños rectángulos R_i, podemos considerar una altura tal que el prisma de base R_i quede (justo) por debajo del gráfico de f.





Así, las sumas inferiores y superiores quedarían definidas como:

$$s(f,\mathscr{P}) = \sum_i A(R_i) \inf_{x\in R_i} f(x),$$

$$S(f,\mathscr{P}) = \sum_i A(R_i) \sup_{x\in R_i} f(x),$$

donde $A(R_i)$ es el área del rectángulo, y por lo tanto el producto $A(R_i)\,\text{ínf}_{x\in R_i} f(x)$ es el volumen del prisma.

Observemos que, una vez que llegamos hasta aquí, el resto ya lo tenemos solucionado. Es decir, las sumas inferiores y superiores son valores reales, y podemos tomar el supremo de todas las sumas inferiores, y el ínfimo de las sumas superiores (como habíamos hecho en una variable), y ver cuándo coinciden.

Ahora, estos conjuntos de sumas inferiores y superiores los formamos variando las particiones entre todas las posibles. ¿Cuáles son las particiones que podemos tomar entonces? Se

nos podría ocurrir particionar D con otras formas que no sean rectángulos. ¿Qué particiones deberían estar permitidas? ¿Y cómo deberíamos definir la medida de estos conjuntos (no rectangulares) que forman la partición?

Dada una función $f : D \to \mathbb{R}$ acotada, con $D \subset \mathbb{R}^2$ acotado, tendríamos que definir:

1. ¿Qué es la medida (el área) de un conjunto en $\mathbb{R}^2$?[1]
2. ¿Qué es una partición?

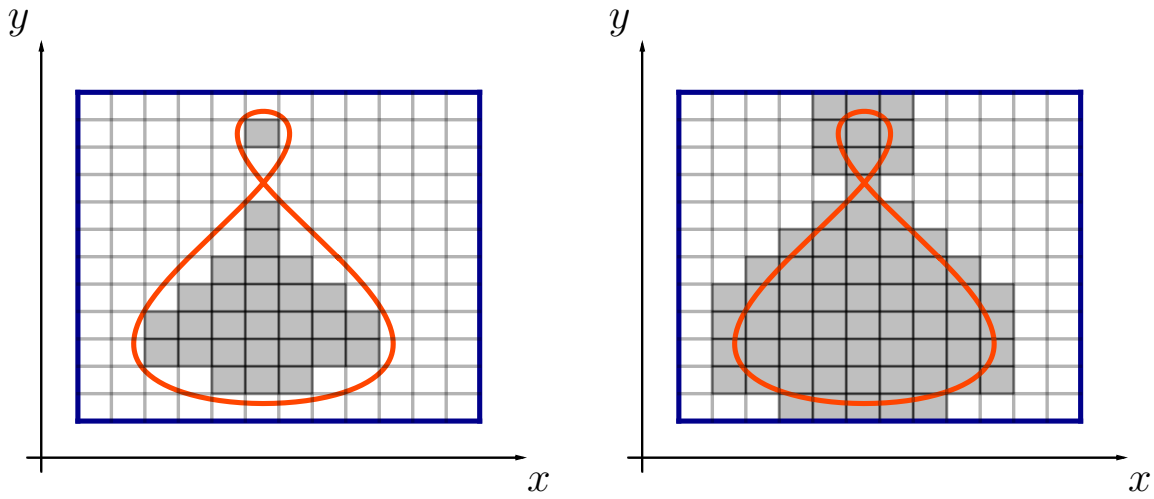
No vamos a hacer un desarrollo detallado de estas dos cuestiones, pero sí daremos algunos lineamientos e ideas básicas, para el caso de dos variables.

Medida de Jordan

Para contestar la primera cuestión, definiremos (muy informalmente) la medida de Jordan. Dado un conjunto $D \subset \mathbb{R}^2$ acotado, describiremos un procedimiento que de alguna forma es similar al de la construcción de la integral de Riemann.

En primer lugar, como D es acotado, está incluido en un rectángulo $[a,b] \times [c,d]$. Partimos $[a,b]$ y $[c,d]$ en n intervalos iguales, que forman un cuadriculado en el rectángulo. Lo que haremos es aproximar el área de D mediante sumas de cuadrados.

Dado un n, y por lo tanto un cuadriculado, llamaremos A_n^- a la suma de las áreas de los cuadrados que quedan completamente incluidos en D, como se ve en la figura (izquierda).



De la misma manera, definimos A_n^+ a la suma de las áreas de los cuadrados que intersecan al conjunto D (es decir, que pueden estar completamente incluidos, o simplemente tocar en

[1] Podríamos pensar que esta pregunta está resuelta a partir del cálculo de integrales de funciones de una variable, donde aprendimos a hallar el área bajo una curva. Sin embargo, esto queda limitado a los conjuntos que son áreas bajo el gráfico de una función de una variable, lo cual es restrictivo.

alguna parte al conjunto D), como se ve a la derecha en la figura. Claramente, tenemos que siempre $A_n^- \leqslant A_n^+$.

A medida que n aumenta, y por lo tanto tenemos un cuadriculado más fino, las aproximaciones A_n^- serán cada vez más grandes, y las aproximaciones A_n^+ serán cada vez más chicas. Es decir, tenemos dos sucesiones de aproximaciones, una creciente, y una decreciente. Además, como tenemos que $A_n^- \leqslant A_n^+$, en particular están acotadas, y por lo tanto son convergentes. Cuando estas sucesiones convergen a un mismo número A, decimos que el conjunto D es medible, y su medida es $m(D) = A$.

Por supuesto, hay conjuntos que no son medibles[2], pero nosotros trabajaremos siempre con conjuntos medibles.

Particiones y definición de integral

Dado un conjunto $\mathscr{D} \subset \mathbb{R}^2$ medible, diremos que una colección de conjuntos $\mathscr{P} = \{D_1, D_2, \ldots, D_n\}$ es una *partición* de D sii

1. D_i es medible Jordan para todo i.
2. $\cup_i D_i = D$
3. $\mathring{D}_i \cap \mathring{D}_j = \emptyset$ para todo $i \neq j$.

Es decir, cada conjunto D_i tiene que ser medible, tienen que cubrir todo el conjunto D, y sus interiores no pueden intersectarse.

Ahora, dado un conjunto $D \subset \mathbb{R}^2$ acotado y medible, $f : D \to \mathbb{R}$ acotada, y $\mathscr{P} = \{D_1, D_2, \ldots, D_n\}$ una partición de D, definimos la suma inferior y superior como:

$$s(f, \mathscr{P}) = \sum_i m(D_i) \inf_{x \in D_i} f(x),$$

$$S(f, \mathscr{P}) = \sum_i m(D_i) \sup_{x \in D_i} f(x).$$

Al igual que en el caso de una variable, cuando

$$\sup_{\mathscr{P}} \Big\{ s(f, \mathscr{P}) \Big\} = \inf_{\mathscr{P}} \Big\{ S(f, \mathscr{P}) \Big\},$$

decimos que f es integrable en D, y a ese número lo denotamos $\iint_D f$.

Por supuesto, esta definición es sumamente incómoda para calcular integrales: deberíamos considerar todas las particiones, y demostrar que el supremo de las sumas inferiores coincide con el ínfimo de las superiores. El camino que estamos siguiendo es el mismo que con integrales

[2] Ejercicio: Comprobar que el conjunto $[0,1] \times [0,1] \cap \mathbb{Q} \times \mathbb{Q}$ no es medible.

de funciones en $\mathbb{R}$: llegar trabajosamente a una definición (que resulta incómoda para demostrar que cierta función es integrable, y más aún calcularla), para luego llegar a teoremas que faciliten los cálculos.

Tenemos entonces, al igual que en cálculo en una variable, que las funciones continuas son integrables en conjuntos compactos (medibles), y los argumentos para demostrarlo son muy similares al caso de una variable.

Trabajaremos entonces a partir de ahora con funciones continuas.

También tenemos propiedades similares a las de la integral en una variable. Listamos las más relevantes a continuación. Las demostraciones siguen la misma lógica, pero no las incluiremos aquí.

1. **Linealidad.** Si f y g son continuas en D, y α y β son reales, entonces
$$\iint_D \alpha f + \beta g = \alpha \iint_D f + \beta \iint_D g\,.$$
2. **Aditividad respecto al dominio.** Si D_1 y D_2 son compactos que no se solapan, y f es continua, entonces
$$\iint_{D_1 \cup D_2} f = \iint_{D_1} f + \iint_{D_2} f\,.$$
3. Si $f \geqslant 0$ en todo el dominio D, entonces $\iint_D f \geqslant 0$.
4. Si $f \geqslant g$ en todo el dominio D, entonces $\iint_D f \geqslant \iint_D g$.
5. $\left|\iint_D f\right| \leqslant \iint_D |f|$.

Sumas de Riemann

Las llamadas *sumas de Riemann* consisten en sumar volúmenes de prismas, pero tomando como altura un valor funcional $f(x_i)$ de algún punto (cualquiera) en el conjunto correspondiente de la partición $x_i \in D_i$. Si llamamos $\mathscr{X}$ a la partición, entonces definimos:

$$SR(f, \mathscr{P}, \mathscr{X}) = \sum_i m(D_i) f(x_i), \quad \text{con } x_i \in D_i.$$

Como cada x_i está en el conjunto D_i correspondiente, entonces su valor funcional $f(x_i)$ está comprendido entre el ínfimo y el supremo:

$$\inf_{x \in D_i} f(x) \leqslant f(x_i) \leqslant \sup_{x \in D_i} f(x).$$

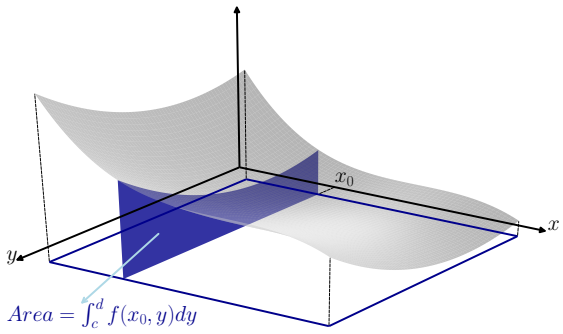


Figura 7.1: Corte para un x_0 fijo. El área coloreada se puede calcular mediante la integral de una variable $\int_c^d f(x_0, y)dy$.

Como esto es cierto para todos los conjuntos D_i de la partición, entonces también es cierta la desigualdad con las sumas. Es decir, cualquier suma de Riemann está comprendida entre las sumas inferior y superior (para la misma partición):

$$s(f, \mathscr{P}) \leqslant SR(f, \mathscr{P}, \mathscr{X}) \leqslant S(f, \mathscr{P}),$$

y por lo tanto si f es continua (y por lo tanto integrable), cualquier suma de Riemann converge al valor de la integral cuando la norma de la partición tiende a cero.

Entonces, si sabemos que f es integrable, esto permite calcular la integral haciendo "una sola cuenta", a diferencia de la definición, con la que debíamos calcular supremos e ínfimos de sumas superiores e inferiores.

Veamos el primer resultado que nos permitirá calcular integrales, utilizando esta observación sobre las sumas de Riemann.

Integrales iteradas

Al inicio de este capítulo vimos una idea intuitiva sobre el cálculo de integrales dobles: aproximar el volumen mediante suma de volúmenes de prismas.

Otra interpretación, haciendo uso de las integrales de funciones de una variable que ya manejamos, es la siguiente.

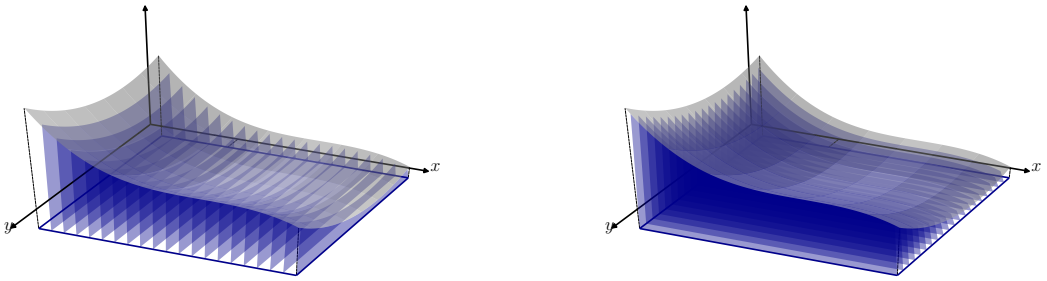
Supongamos que tenemos una función definida en un rectángulo $[a,b] \times [c,d]$. Consideremos un valor fijo de x en $[a,b]$, digamos x_0, y consideremos la función (de una variable) que resulta de f restringiendo la primera variable a x_0, es decir, la función $f(x_0, y)$.

Entonces podemos calcular la integral entre c y d de la función $f(x_0, y)$, y esto nos da el área pintada en la figura. La idea ahora es "sumar" todas estas áreas, para cada valor de x (que habíamos fijado en x_0), entre a y b.

Esto corresponde a integrar el área $A(x)$ para cada x entre a y b. Es decir, según esta intuición geométrica, la integral de f sobre el rectángulo $D = [a,b] \times [c,d]$ resultaría:

$$\iint_D f = \int_a^b A(x)dx = \int_a^b \underbrace{\left(\int_c^d f(x,y)dy\right)}_{A(x)} dx.$$

Por otro lado, parece lógico que deberíamos llegar al mismo resultado si realizamos el mismo procedimiento, pero realizando cortes en el otro sentido, como indica la figura.



El resultado que veremos a continuación establece dos elementos importantes: en primer lugar, que la integral que definimos arriba, se puede calcular efectivamente como una integral iterada. En segundo lugar, que podemos intercambiar el orden de integración, obteniendo el mismo resultado.

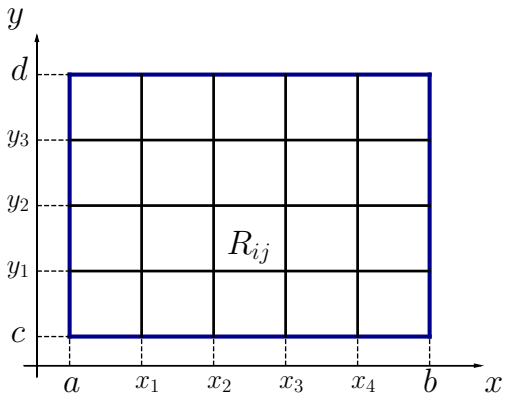
Teorema 7.1 (Fubini, Versión I)**.** Sea $f : [a,b] \times [c,d] \to \mathbb{R}$ continua. Entonces:

$$\iint_D f = \int_a^b \int_c^d f(x,y)dydx = \int_c^d \int_a^b f(x,y)dxdy.$$

Observación 7.2
La primera igualdad establece que podemos calcular la integral definida como una integral iterada. La segunda igualdad establece que, realizando la integral iterada en otro orden, el resultado es el mismo.

Demostración. Vamos a demostrar solamente la primera igualdad. La segunda se prueba de forma análoga.

Consideremos las particiones $\mathscr{P}_1 = \{x_0, x_1, \ldots, x_n\}$ de $[a,b]$, y $\mathscr{P}_2 = \{y_0, y_1, \ldots, y_m\}$ de $[c,d]$. Esto define también una partición $\mathscr{P}$ del rectángulo $[a,b] \times [c,d]$, mediante los rectángulos que llamaremos $R_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$.



Vamos a llamar $\varphi(x)$ a la función que a cada x le asigna el área de la sección indicada en la Figura 7.1. Esto es, para cada x fijo, tenemos $\varphi(x) = \int_c^d f(x,y)dy$. Como observamos antes, esta función es la que integraremos para obtener el volumen. Es decir:

$$\iint_D f = \int_a^b \underbrace{\int_c^d f(x,y)dy}_{\varphi(x)} dx$$

Tenemos entonces dos funciones, $f(x,y)$ (de dos variables) definida en $[a,b] \times [c,d]$ y $\varphi(x)$ (de una variable) definida en $[a,b]$.

Vamos a demostrar que

$$s(f,\mathscr{P}) \leqslant s(\varphi,\mathscr{P}_1) \leqslant S(\varphi,\mathscr{P}_1) \leqslant S(f,\mathscr{P}).$$

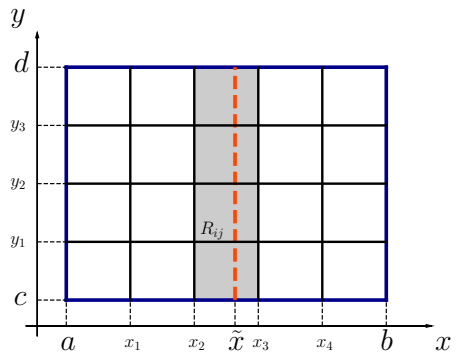
Observar que las sumas de los extremos (de f) son las que definimos en este capítulo, mientras que las dos sumas del medio (de φ) son las sumas definidas en el curso de cálculo en una variable.

Por otro lado, observemos que esta cadena de desigualdades prueba lo que buscamos, ya que como sabemos que f es integrable, podemos hacer las sumas inferiores y superiores de f tan cerca como queramos. Como estas sumas son los extremos de la cadena de desigualdades, las sumas de φ que quedan atrapadas, también convergen a la integral.

Demostremos entonces la primera desigualdad: $s(f,\mathscr{P}) \leqslant s(\varphi,\mathscr{P}_1)$. La segunda es trivial, a partir de las definiciones, y la última se demuestra de forma análoga a la primera.

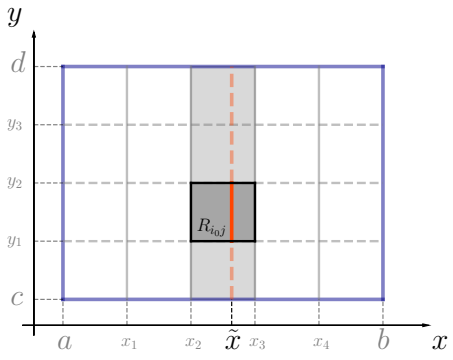
Como f es continua, el ínfimo de los valores funcionales en cada rectángulo, es en realidad un mínimo. Entonces la suma inferior es

$$s(f,\mathscr{P}) = \sum_{i=0}^{n}\sum_{j=0}^{m} m(R_{ij}) \min_{R_{ij}} f(x,y) = \sum_{i=0}^{n}\sum_{j=0}^{m} (x_{i+1}-x_i)(y_{j+1}-y_j) \min_{R_{ij}} f(x,y).$$



Ahora, consideremos un valor $\tilde{x} \in [x_{i_0}, x_{i_0+1}]$, y llamemos $g(y)$ a la función que resulta de restringir f a $\tilde{x}$ en su primera coordenada: $g(y) = f(\tilde{x}, y)$, que en la figura corresponde a evaluar f en el segmento vertical indicado. Trabajaremos ahora con los rectángulos de la columna que contiene dicho segmento (esto corresponde a fijar el índice i_0), que en la figura están sombreados.

Para cada uno de los rectángulos sombreados, digamos $R_{i_0 j}$, tenemos que el mínimo de $f(x,y)$ en todo ese rectángulo es menor o igual que el mínimo de $f(\tilde{x},y)$ en el segmento señalado que queda incluido en $R_{i_0 j}$. Esto es claro porque estamos tomando el mínimo en una región más grande (el rectángulo), que contiene al segmento. Tenemos entonces que



$$\min_{(x,y)\in R_{i_0 j}} f(x,y) \leqslant \min_{y\in[y_j,y_{j+1}]} f(\tilde{x},y)$$

Como esto es cierto para cada uno de los rectángulos de la columna sombreada, podemos sumar todos estos mínimos, multiplicados por el alto correspondiente a cada rectángulo. Recordando que habíamos llamado $g(y) = f(\tilde{x},y)$, esto es:

$$\sum_j (y_{j+1} - y_j) \min_{(x,y)\in R_{i_0 j}} f(x,y) \leqslant \sum_j (y_{j+1} - y_j) \min_{y\in[y_j,y_{j+1}]} g(y).$$

Ahora, el lado derecho de esta desigualdad es una suma inferior para g con la partición $\mathscr{P}_2$, y por lo tanto (por ser una suma inferior) es menor a la integral correspondiente. Es decir,

$$\sum_j (y_{j+1} - y_j) \min_{y\in[y_j,y_{j+1}]} g(y) = s(g,\mathscr{P}_2) \leqslant \int_c^d g(y)dy = \int_c^d f(\tilde{x},y)dy = \varphi(\tilde{x}).$$

En definitiva, como esto lo hicimos para un $\tilde{x}$ cualquiera dentro del intervalo $[x_{i_0}, x_{i_0+1}]$, tenemos que

$$\sum_j (y_{j+1} - y_j) \min_{(x,y)\in R_{i_0 j}} f(x,y) \leqslant \varphi(\tilde{x}), \quad \forall \tilde{x} \in [x_{i_0}, x_{i_0+1}].$$

Y como esta desigualdad es cierta para todo $\tilde{x}$ en el intervalo, en particular también es cierta para el mínimo de φ en el intervalo:

$$\sum_j (y_{j+1} - y_j) \min_{(x,y)\in R_{i_0 j}} f(x,y) \leqslant \min_{\tilde{x}\in[x_{i_0},x_{i_0+1}]} \varphi(\tilde{x}).$$

Para que aparezca el área del rectángulo Ri_0j, multiplicamos a ambos lados por el ancho $(x_{i_0+1} - x_{i_0})$:

$$(x_{i_0+1} - x_{i_0}) \sum_j (y_{j+1} - y_j) \min_{(x,y)\in R_{i_0 j}} f(x,y) \leqslant (x_{i_0+1} - x_{i_0}) \min_{\tilde{x}\in[x_{i_0}, x_{i_0+1}]} \varphi(\tilde{x}).$$

A esta desigualdad llegamos para la columna sombreada, pero por supuesto es cierta para toda columna, por lo tanto podemos sumar las desigualdades correspondientes a cada columna (esto es, sumar en i), llegando a

$$\sum_i (x_{i_0+1} - x_{i_0}) \sum_j (y_{j+1} - y_j) \min_{(x,y)\in R_{i_0 j}} f(x,y) \leqslant \sum_i (x_{i_0+1} - x_{i_0}) \min_{\tilde{x}\in[x_{i_0}, x_{i_0+1}]} \varphi(\tilde{x}).$$

Observar que el lado izquierdo es exactamente la suma inferior de f en $\mathscr{P}$, y el lado derecho es la suma inferior de φ en $\mathscr{P}_1$, que era lo que queríamos probar.

$$\underbrace{\sum_i \sum_j (x_{i_0+1} - x_{i_0})(y_{j+1} - y_j) \min_{(x,y)\in R_{i_0 j}} f(x,y)}_{s(f,\mathscr{P})} \leqslant \underbrace{\sum_i (x_{i_0+1} - x_{i_0}) \min_{\tilde{x}\in[x_{i_0}, x_{i_0+1}]} \varphi(\tilde{x})}_{s(\varphi,\mathscr{P}_1)}.$$

□

Sobre la notación: Observar que los símbolos de integral y sus correspondientes dx se leen "desde afuera hacia adentro" (la integral de la izquierda se corresponde con el diferencial de la derecha). Además, hay que interpretar esto como se describió más arriba: *para cada x entre a y b, integramos la función $f(x,y)$ variando y entre c y d.*

$$\int_a^b \int_c^d f(x,y)\,dy\,dx$$

Calculemos entonces, utilizando este resultado, la primera integral doble.

Ejemplo 7.3
Calculemos la integral de la función $f(x,y) = x^2y^3$ en el dominio $D = [0,1] \times [1,2]$.

$$\iint_D f(x,y) = \int_0^1 \int_1^2 x^2y^3\,dy\,dx.$$

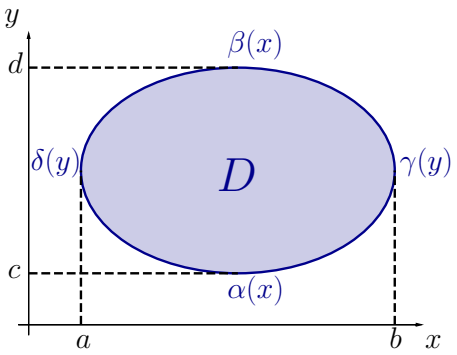
Hay que integrar entonces la función x^2y^3 respecto de la variable y. Por lo tanto, para esta primera integral, consideramos x como constante, y resulta

$$\int_0^1 \int_1^2 x^2y^3\,dy\,dx = \int_0^1 x^2 \cdot \frac{y^4}{4}\Big|_{y=1}^{y=2} dx = \int_0^1 x^2 \frac{(2^4 - 1^4)}{4} dx = \frac{15}{4} \cdot \frac{x^3}{3}\Big|_{x=0}^{x=1} = \frac{15}{12}.$$

En este caso es inmediato corroborar que la integral planteada en el otro sentido $\int_1^2 \int_0^1 x^2 y^3 dxdy$ da el mismo resultado.

Si bien este teorema es sumamente útil, ya que nos permitió calcular la primera integral doble, y además asegura que podemos recorrer el dominio en los dos sentidos para hallar la integral, nos limita los dominios de integración a rectángulos. Veamos una segunda versión del Teorema de Fubini, aunque con un enunciado muy poco formal.

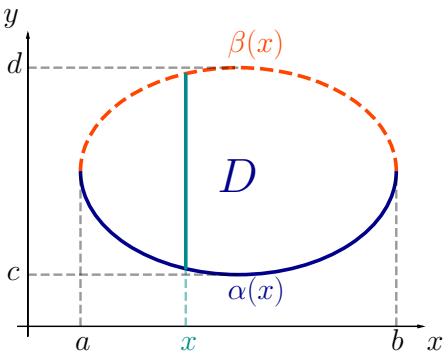
Teorema 7.4 (Fubini, Versión II)**.** Sea $f : D \to \mathbb{R}$ continua, y D como en la figura[3].

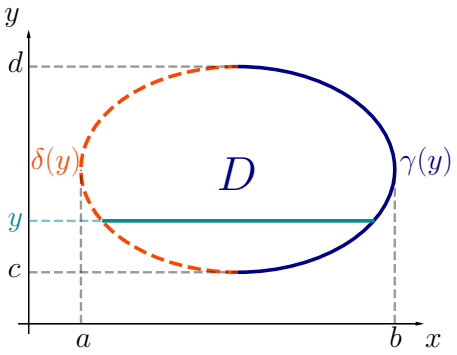


Entonces:

$$\iint_D f = \int_a^b \int_{\alpha(x)}^{\beta(x)} f(x,y)dydx = \int_c^d \int_{\delta(y)}^{\gamma(y)} f(x,y)dxdy.$$

Es decir, la expresión $\int_a^b \int_{\alpha(x)}^{\beta(x)} f(x,y)dydx$ la interpretamos de la siguiente forma: para cada valor de x entre a y b, variamos y desde el valor $\alpha(x)$ (el "piso" del segmento vertical en el dibujo) hasta el valor $\beta(x)$ (el "techo" del segmento).





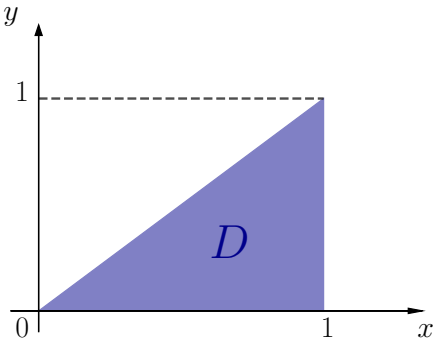
Del mismo modo, la expresión $\int_c^d \int_{\delta(y)}^{\gamma(y)} f(x,y)dxdy$ la interpretamos así: para cada valor de y (indicado en la integral de "afuera") entre c y d, que son los valores máximos donde puede variar y, recorremos el dominio de forma horizontal, variando x desde el valor $\delta(y)$ hasta el valor $\gamma(y)$.

[3]No necesariamente una elipse, por supuesto, sino que se pueda expresar como la región comprendida entre dos funciones, en ambos sentidos.

Observación 7.5
- *Si el dominio no se puede describir como en la figura, se puede descomponer como unión de dominios más simples (y por lo tanto suma de integrales en esos dominios).*
- *La integral "de afuera" **siempre** debe tener constantes en los límites de integración.*

Cuando integramos funciones de dos variables entonces, tenemos dos formas de recorrer el dominio, y por lo tanto tenemos que elegir una de ellas para realizar los cálculos. ¿Será que algún orden de integración resulta más conveniente que otro? Veamos un ejemplo.

Ejemplo 7.6
Calculemos la integral de $f(x,y) = e^{-x^2}$ en el dominio de la figura.



Planteando las dos formas de recorrer el dominio, tenemos que podemos calcular la integral de estas dos formas:

$$\iint_D f = \int_0^1 \int_0^x e^{-x^2} dydx = \int_0^1 \int_y^1 e^{-x^2} dxdy.$$

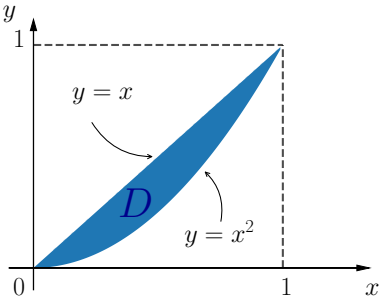
Comencemos por la última: $\int_0^1 \int_y^1 e^{-x^2} dxdy$. Observemos que lo primero que hay que resolver es la integral del medio, respecto a la variable x, de la función e^{-x^2}. Pero esta función no tiene primitiva elemental, por lo que no vamos a poder avanzar más siguiendo este camino.

Por otro lado, si planteamos la integral según el otro orden de integración, obtenemos:

$$\int_0^1 \int_0^x e^{-x^2} dydx = \int_0^1 e^{-x^2} y \Big|_{y=0}^{y=x} dx = \int_0^1 xe^{-x^2} dx = \frac{1}{2}\left(1 - \frac{1}{e}\right).$$

Ejercicio 7.7
Veamos ahora cómo plantear las dos formas de recorrer el siguiente dominio, para integrar una función genérica $f(x,y)$. Corroborar que:

$$\iint_D f = \int_0^1 \int_{x^2}^x f(x,y) dydx = \int_0^1 \int_y^{\sqrt{y}} f(x,y) dxdy.$$



Observación 7.8
Si integramos la función de una variable constante uno $f(x) = 1$, en un intervalo $[a,b]$,

obtenemos como resultado la longitud del intervalo $(b-a)$. A partir de la definición, es fácil ver que sucede lo mismo para funciones definidas en el plano: si integramos $f(x,y)=1$ en un conjunto D, lo que obtenemos es el área del conjunto: $\iint_D 1dxdy = Area(D)$.

Finalicemos esta sección con un ejercicio que brinda otra demostración de un resultado del capítulo pasado.

Ejercicio 7.9
Usar el Teorema de Fubini para dar una demostración alternativa al Teorema 6.23 (de Schwarz, o de las derivadas cruzadas).
Sugerencia: suponer que $f_{xy}(x_0,y_0)-f_{yx}(x_0,y_0)>0$. Como son continuas por hipótesis, existe un rectángulo donde $f_{xy}(x,y)-f_{yx}(x,y)>0$ para todo punto (x,y) del rectángulo.

Ejemplo 7.10
Si bien en la próxima sección veremos el contexto y varios ejemplos de cambios de variables en $\mathbb{R}^n$, muchas veces también necesitaremos hacer cambios de variable de cálculo en una variable, para hallar primitivas intermedias. Tomemos por ejemplo esta integral doble:

$$\int_0^1 \int_0^{1-x} \frac{1}{(1+x+y)^2} dydx.$$

Entonces para resolver la primera integral, $\int_0^{1-x} \frac{1}{(1+x+y)^2} dy$, debemos calcular una primitiva de $\frac{1}{(1+x+y)^2}$ respecto de y (pensando en x como una constante por ahora). Podemos hacer entonces un cambio de variable de los que estamos acostumbrados de cursos anteriores, como $u=1+x+y$. De nuevo, aquí pensemos en x como una constante, ya que estamos integrando respecto de la variable y. Una primitiva es entonces:

$$\int \frac{1}{(1+x+y)^2} dy = \int \frac{1}{u^2} du = \frac{-1}{u} = \frac{-1}{1+x+y}.$$

Volviendo a la integral doble y evaluando,

$$-\int_0^1 \frac{1}{1+x+y}\Big|_{y=0}^{y=1-x} dx = -\int_0^1 \frac{1}{1+x+1-x} - \frac{1}{1+x} dx = -\int_0^1 \frac{1}{2} dx + \int_0^1 \frac{1}{1+x} dx.$$

Luego, el resultado es

$$\int_0^1 \int_0^{1-x} \frac{1}{(1+x+y)^2} dydx = \ln(2) - \frac{1}{2}.$$

7.1. Cambios de Variable

Al igual que para funciones de una variable, muchas veces es conveniente realizar un cambio de variable para calcular integrales. En este contexto de integrales múltiples, vamos a tener que considerar con cuidado cómo se modifica el dominio de integración, además de la función a integrar.

Las funciones que podemos usar como cambio de variable deben cumplir algunas condiciones técnicas, que describimos a continuación.

Definición 7.11. Sean U y V dos abiertos de $\mathbb{R}^2$. Decimos que $g : U \to V$ es un cambio de variable (o un difeomorfismo) sii g es biyectiva, diferenciable, y $\det(J_g) \neq 0$ en todo el dominio U.

Se le llama difeomorfismo a una función diferenciable con inversa diferenciable. En la definición, como g es biyectiva tiene inversa, y la condición sobre el determinante de la matriz Jacobiana garantiza que la inversa es diferenciable, gracias al Teorema de la Función Inversa, que se verá en un curso posterior.

Para ganar intuición sobre cómo funcionan los cambios de variable en las integrales múltiples, comencemos trabajando con las funciones más sencillas: las lineales.

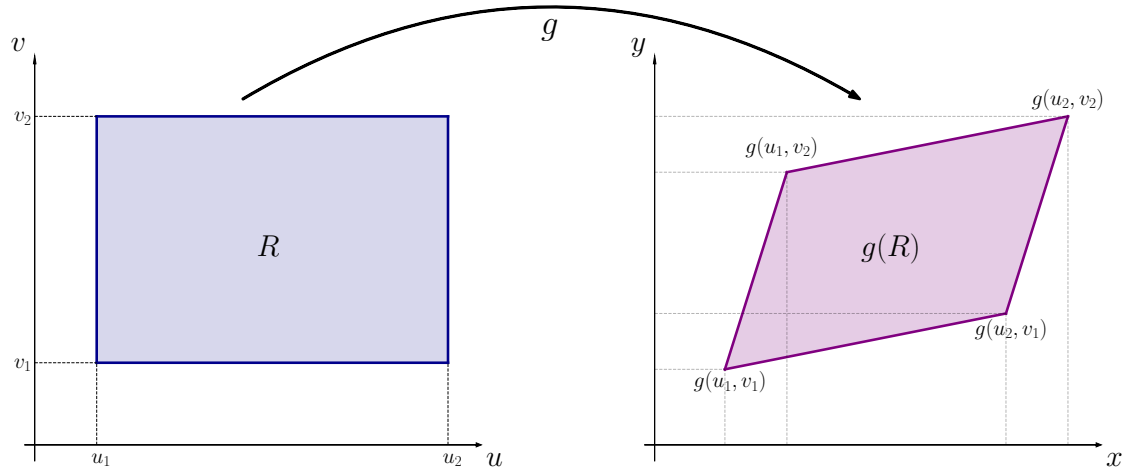
Un cambio de variable lineal (de $\mathbb{R}^2$ a $\mathbb{R}^2$) es una función $g(u,v) = (Au + Bv, Cu + Dv)$. Calculemos la matriz Jacobiana, para asegurarnos de estar en las condiciones de la definición.
Derivando, obtenemos que la matriz es: $J_g(u,v) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. El determinante es $\det(J_g) = AD - BC$, por lo que debemos pedir que $AD \neq BC$.

En realidad...

... ya sabíamos que esa debía ser la matriz Jacobiana, pues ¿cuál es la transformación lineal que mejor aproxima una transformación lineal?
Ella misma, por supuesto. Y la matriz obtenida no es otra que la matriz asociada a la transformación lineal g. Además, de los cursos de álgebra lineal, ya sabíamos que para que una transformación lineal sea invertible, el determinante de su matriz asociada debe ser no nulo.

Como comenzamos los análisis sobre integrales utilizando rectángulos (tanto para el dominio como para los elementos de las particiones), veamos cómo se transforman los rectángulos al aplicarles g.

Consideremos un rectángulo como en la figura (izquierda), de vértices $M = (u_1, v_1)$, $N = (u_2, v_1)$, $P = (u_1, v_2)$, y $Q = (u_2, v_2)$. A la derecha tenemos la imagen del rectángulo por g, que es un paralelogramo de vértices $M' = g(u_1, v_1)$, $N' = g(u_2, v_1)$, etc. Lo que nos interesa de estos rectángulos (o paralelogramos) son sus áreas. El área del rectángulo R es trivialmente $A(R) = (u_2 - u_1)(v_2 - v_1)$.



Por otro lado, el área del paralelogramo $g(R)$ se puede calcular como el módulo del producto vectorial de los lados $(N'-M')$ y $(P'-M')$:

$$\begin{aligned} A(g(R)) &= \|(N'-M') \wedge (P'-M')\| = \|(g(u_2,v_1)-g(u_1,v_1)) \wedge (g(u_1,v_2)-g(u_1,v_1))\| \\ &= \| \left(A(u_2-u_1), C(u_2-u_1)\right) \wedge \left(B(v_2-v_1), D(v_2-v_1)\right) \| \\ &= \underbrace{(u_2-u_1)(v_2-v_1)}_{A(R)} \cdot \underbrace{\|(A,C) \wedge (B,D)\|}_{|det(J_g)|}. \end{aligned}$$

Es decir, una transformación lineal modifica el área de un rectángulo,[4] multiplicándola por el valor absoluto del determinante de su matriz asociada:

$$A\left(g(R)\right) = A(R) \cdot |det(J_g)|.$$

Observar que si el determinante es cero, el área de la imagen es cero. Esto tiene sentido, pues en ese caso la transformación lineal colapsa todo en una dimensión menor (en una recta o en un punto).

A partir de esta relación entre las áreas, veamos qué sucede con la integral de una función continua.

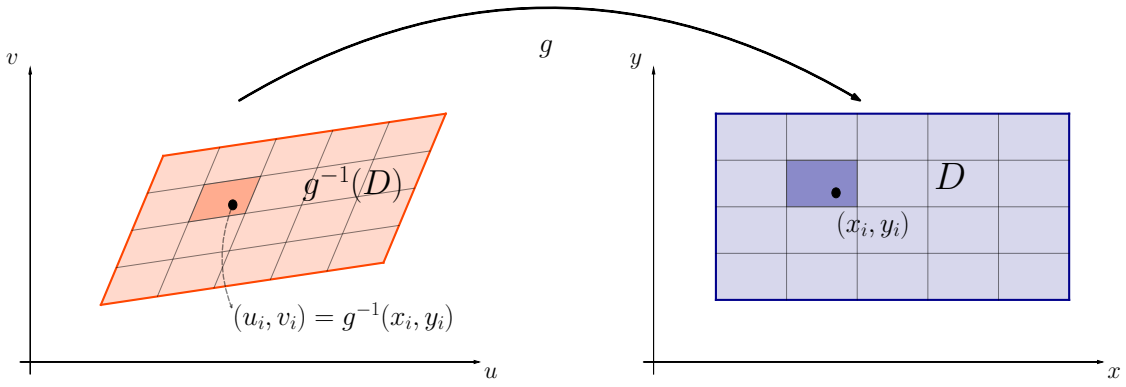
Consideremos una función continua f definida en un dominio rectangular $D \subset \mathbb{R}^2$, y una partición en rectángulos R_i, como antes. Entonces cualquier suma de Riemann converge a la integral, a medida que achicamos los rectángulos de la partición:

$$SR(f, \mathscr{P}) = \sum_i A(R_i) f(x_i, y_i) \longrightarrow \iint_D f(x,y),$$

donde los puntos (x_i, y_i) donde se evalúa la función, están en los rectángulos correspondientes $(x_i, y_i) \in R_i$.

[4] En realidad de cualquier conjunto.

Supongamos además que tenemos una transformación lineal g, que tiene como **codominio** el plano $\mathbb{R}^2$ donde tenemos definida la función f. Es decir, el dominio de f (el rectángulo que llamamos D), es la imagen por g de algún paralelogramo.[5] Vamos a mirar todos los elementos de la suma de Riemann, pero en "el otro lado", es decir, vamos a considerar sus preimágenes por g.



Por ejemplo, la preimagen por g del rectángulo D (o lo que es lo mismo, la imagen por su inversa g^{-1}), es $g^{-1}(D)$, que es un paralelogramo (pues g^{-1} es también una transformación lineal).

Las preimágenes de los elementos de la partición R_i, serán también paralelogramos $g^{-1}(R_i)$, que denominamos $P_i = g^{-1}(R_i)$. En cada rectángulo R_i tenemos los puntos (x_i, y_i), por lo tanto dentro de cada paralelogramo P_i tenemos los puntos $g^{-1}(x_i, y_i)$, que denominaremos $(u_i, v_i) = g^{-1}(x_i, y_i)$.

Además, sabemos cómo se relacionan las áreas de los elementos de las particiones: el área del rectángulo R_i es exactamente el área del paralelogramo P_i multiplicada por el valor aboluto del determinante de la matriz Jacobiana de g. Esto es: $A(R_i) = A(P_i)|\det(J_g)|$.

En la suma de Riemann de f en D, teníamos una suma de áreas de rectángulos R_i por valores funcionales en (x_i, y_i). Ahora relacionamos esas áreas de los rectángulos R_i con áreas de paralelogramos P_i, y los puntos (x_i, y_i) los escribimos como $(x_i, y_i) = g(u_i, v_i)$. Tenemos entonces que

$$SR(f, \mathscr{P}) = \sum_i \underbrace{A(R_i)}_{A(P_i)|\det(J_g)|} \; f\underbrace{(x_i, y_i)}_{g(u_i, v_i)} = \sum_i A(P_i) \cdot |\det(J_g)| f\left(g(u_i, v_i)\right).$$

El término de la derecha es también una suma de Riemann, pero de otra función, en otro dominio (pues son todos elementos de "la izquierda" en la figura). Como los paralelogramos

[5] Recordemos que la inversa de g, que denominamos g^{-1}, es también lineal.

P_i forman una partición de $g^{-1}(D)$, lo que tenemos es la suma de Riemann de la función $|\det(J_g)|f(g(u_i,v_i))$ en $g^{-1}(D)$.

Como la función $(f\circ g)$ es continua, entonces esta suma de Riemann converge a la integral. Tenemos entonces que:

$$\begin{array}{ccc} \sum_i A(R_i)f(x_i,y_i) & = & \sum_i A(P_i)\cdot|\det(J_g)|f(g(u_i,v_i)) \\ \downarrow & & \downarrow \\ \iint_D f(x,y)dxdy & & \iint_{g^{-1}(D)} f(g(u,v))\,|\det(J_g)|dudv \end{array}$$

Y, como el límite debe ser único, llegamos a que:

$$\iint_D f(x,y)dxdy = \iint_{g^{-1}(D)} f(g(u,v))\,|\det(J_g)|dudv.$$

Es decir, podemos calcular la integral de f en D, como la integral de una **nueva función** (que es $f\circ g$ multiplicada por el determinante de la Jacobiana) en un **nuevo dominio**, que es $g^{-1}(D)$.

Se puede ver que esta expresión es también cierta cuando g es un difeomorfismo (no necesariamente una transformación lineal), y el argumento principal es que localmente, como g es diferenciable, podemos aproximar muy bien a g por su diferencial, que es una transformación lineal. El resultado es que obtenemos la misma fórmula, pero la matriz Jacobiana tiene que estar evaluada en cada punto (u,v), que es donde se realiza la aproximación.

Con esto entonces tenemos el siguiente resultado.

Teorema 7.12 (Cambio de Variable)**.** Sea $f:D\to\mathbb{R}$ continua, y $g:U\to V$ un difeomorfismo, donde U y V son abiertos, y $D\subset V$. Entonces,

$$\iint_D f(x,y)dxdy = \iint_{g^{-1}(D)} f(g(u,v))\,|\det(J_g(u,v))|dudv.$$

Hay tres elementos a tener en cuenta a la hora de realizar un cambio de variables.
El primero es **no olvidarse del determinante de la matriz Jacobiana**.
Los otros dos están relacionados a cómo elegir el cambio de variables, ya que cuando se elige una función g de cambio de variables, se modifican **dos cosas**:

- **La función** a integrar, por supuesto.
- **El dominio**. La forma del dominio puede cambiar mucho: recordar que luego del cambio de variable hay que integrar en $g^{-1}(D)$. Cuando realizamos cambios de variable en integrales de funciones en $\mathbb{R}$, cambia también el dominio, pues hay que modificar los límites de integración. Pero siempre terminamos integrando en un intervalo.

Veamos primero un ejemplo sencillo, con un cambio de variable lineal, para luego sí analizar ejemplos más complejos.

Ejemplo 7.13
Calculemos la integral de $f(x,y) = e^{\frac{x-y}{x+y}}$ *en el dominio definido por* $D = \{(x,y) \in \mathbb{R}^2 : x \geqslant 0, y \geqslant 0, x+y \leqslant 1\}$.

En este caso, tenemos un dominio sencillo, pero la función presenta complicaciones en el exponente a la hora de integrar. Si realizamos un cambio de variable lineal: $u = x+y$, $v = x-y$, *entonces la función resulta* $e^{v/u}$, *que parece más sencilla para integrar. Veamos con cuidado cómo son los elementos de este cambio de variable.*

Para empezar, veamos cuál es la función de cambio de variable y su inversa. Observar que la función g tiene que expresar las variables "viejas" x e y, en función de las variables "nuevas" u y v.

De la elección conveniente que hicimos del cambio de variable, podemos despejar x e y:

$$\begin{aligned} u &= x+y \\ v &= x-y \end{aligned} \qquad \Longrightarrow \qquad \begin{aligned} x &= \frac{u+v}{2} \\ y &= \frac{u-v}{2} \end{aligned}$$

Es decir, la función $g(u,v)$ *es* $g(u,v) = \left(\frac{u+v}{2}, \frac{u-v}{2}\right)$, *y su inversa es* $g^{-1}(x,y) = (x+y, x-y)$.

Veamos entonces los tres elementos que necesitamos para plantear la integral con el cambio de variable: el nuevo dominio, la función (componiendo f y g), y el Jacobiano.

Comencemos por la función, que es lo más sencillo. La nueva función es

$$f(g(u,v)) = f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) = e^{\frac{\frac{u+v}{2}-\frac{u-v}{2}}{\frac{u+v}{2}+\frac{u-v}{2}}} = e^{\frac{v}{u}}.$$

En realidad, como ***elegimos*** *el cambio de forma que $u = x+y$, y $v = x-y$, ya sabíamos que el resultado era ese.*

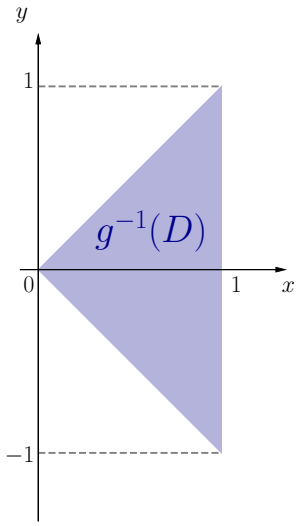
Calculemos ahora la matriz Jacobiana de g. Tenemos que $J_g(u,v) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$, *por lo tanto el valor absoluto del determinante es* $|J_g(u,v)| = 1/2$.

Finalmente, el nuevo dominio de integración es $g^{-1}(D)$. En este caso, como el dominio original D es el interior de un triángulo, y el cambio de variable es una transformación lineal (y por lo tanto su inversa también es lineal), entonces la imagen de D por g^{-1} también será el interior de un triángulo. Como consecuencia, es suficiente calcular las preimágenes de los vértices por g.

Tenemos:

$$g^{-1}(0,0) = (0,0)$$
$$g^{-1}(1,0) = (1,1)$$
$$g^{-1}(0,1) = (1,-1).$$

Por lo tanto el nuevo dominio es el de la figura.



En este caso también podemos observar que las tres desigualdades que definen D, las podemos re-escribir en términos de las nuevas variables:

$$\begin{aligned} x \geqslant 0 &\implies \frac{u+v}{2} \geqslant 0 \implies u \geqslant -v, \\ y \geqslant 0 &\implies \frac{u-v}{2} \geqslant 0 \implies u \geqslant v, \\ x+y \leqslant 1 &\implies u \leqslant 1. \end{aligned}$$

Entonces llegamos a otra descripción del nuevo dominio, que es $g^{-1}(D) = \{(u,v) \in \mathbb{R}^2 : u \geqslant -v, u \geqslant v, u \leqslant 1\}$.

Uniendo estos elementos, el cambio de variable para esta integral queda:

$$\iint_D f(x,y)dxdy = \iint_{g^{-1}(D)} f(g(u,v)) \,|\det(J_g(u,v))|\, dudv = \iint_{g^{-1}(D)} e^{\frac{v}{u}} \cdot \frac{1}{2} dudv.$$

Una forma de recorrer este dominio es la siguiente: para cada valor de u entre 0 *y* 1*, recorremos verticalmente con v entre el "piso" ($v = -u$) hasta el "techo" ($v = u$).*

$$\int_0^1 \int_{-u}^{u} e^{\frac{v}{u}} \cdot \frac{1}{2} dv du.$$

El resto del cálculo queda como ejercicio, ya que es una integral doble sencilla.

7.1.1. Coordenadas Polares

Un cambio de variable muy útil para calcular algunas integrales dobles, es el cambio a polares. Estas coordenadas en realidad ya son muy familiares: las usamos con los números complejos en el primer capítulo, y luego fueron una herramienta importante para calcular límites en el Capítulo 5.

Recordemos que podemos representar cada punto del plano mediante su distancia al origen ρ y el ángulo θ que forma con el eje horizontal. Analíticamente, tenemos:

$$\begin{aligned} x &= \rho \cos(\theta), \\ y &= \rho \sin(\theta). \end{aligned}$$

En términos de la notación que estamos usando para los cambios de variable[6], tenemos $g : [0,\infty) \times [0,2\pi) \to \mathbb{R}^2$, con $g(\rho,\theta) = (\rho\cos(\theta), \rho\sin(\theta))$. Calculemos el determinante de la matriz Jacobiana:

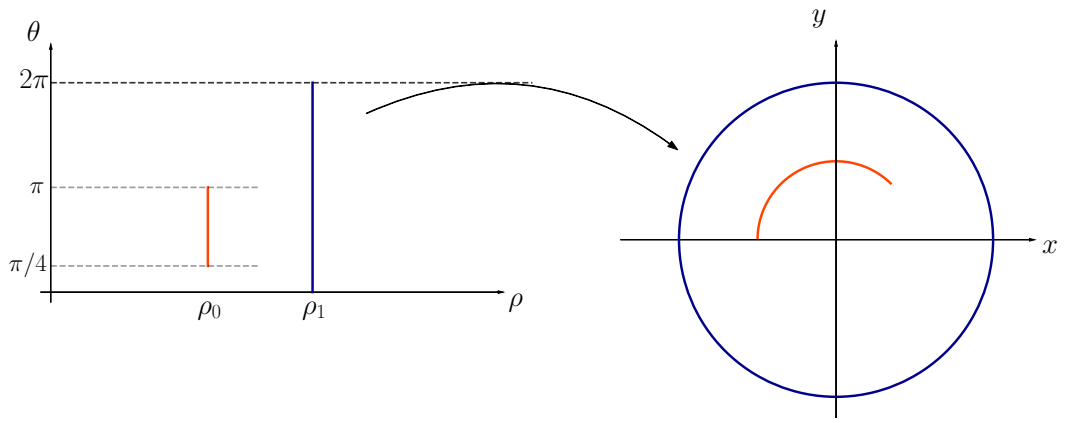
$$J_g(\rho,\theta) = \begin{pmatrix} \cos(\theta) & -\rho\sin(\theta) \\ \sin(\theta) & \rho\cos(\theta) \end{pmatrix},$$

de donde tenemos $|det(J_g(\rho,\theta))| = |\rho\cos^2(\theta) + \rho\sin^2(\theta)| = \rho$.

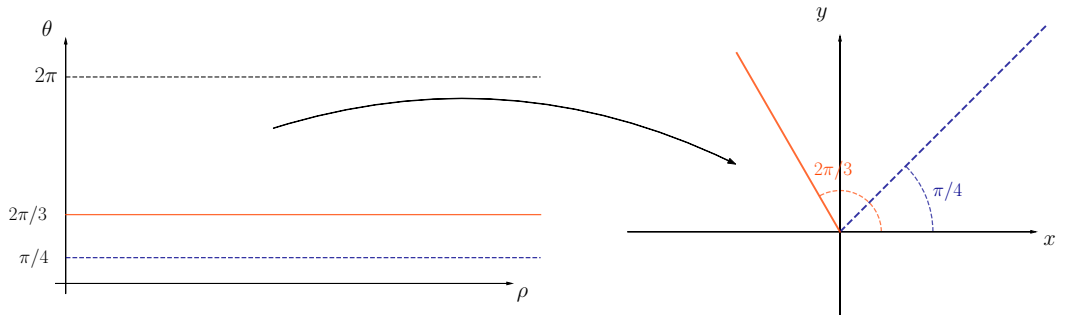
Para ganar intuición sobre cómo se comporta g, veamos cómo transforma algunos conjuntos particulares.

Por ejemplo, un segmento vertical $\{\rho_0\} \times [0,2\pi)$ como el de la figura, se transforma por g en una circunferencia de radio ρ_0. Si tomamos un segmento que no cubre los 2π, como por ejemplo $\{\rho_1\} \times [\pi/4, \pi)$, obtenemos un arco en vez de la circunferencia entera.

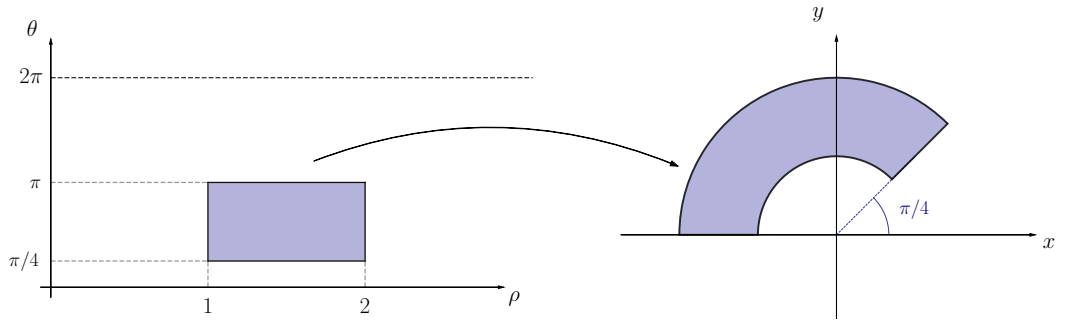
[6] En este caso, en vez de utilizar u y v para las nuevas variables, usamos ρ y θ.



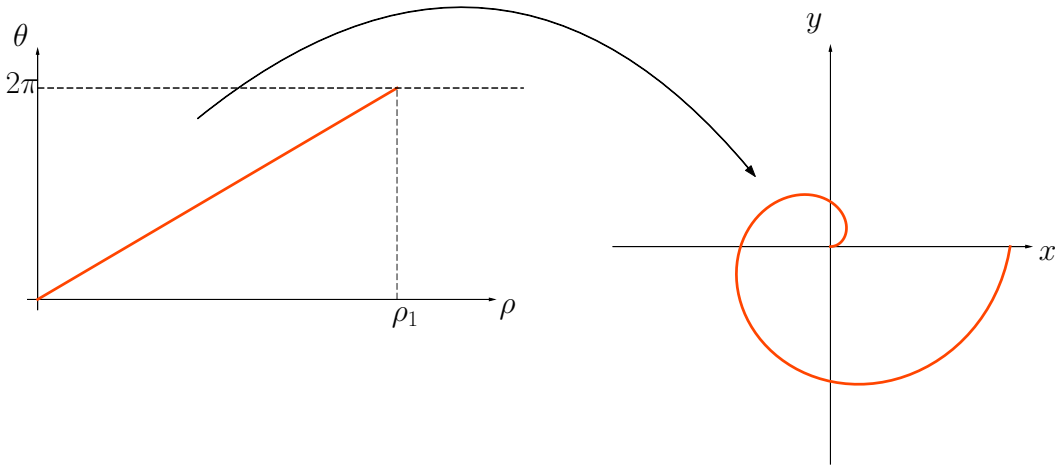
Los segmentos con θ_0 fijo, y ρ variando entre 0 y ∞ por ejemplo, se transforman en semi-rectas que parten del origen, y forman un ángulo de θ_0 con el eje horizontal.



Juntando ambas ideas, un rectángulo como $[1,2]\times[\pi/4,\pi]$ se transforma en una parte de un anillo.



Finalmente, si variamos tanto ρ como θ según una recta como la de la figura, obtenemos una curva que, a medida que va aumentando el ángulo, también se va alejando del origen.



Volviendo a las integrales dobles, veamos un ejemplo donde el cambio de variables a polares facilita sensiblemente las cuentas.

Ejemplo 7.14
Calculemos $\iint_D f(x,y)$ *donde* $f(x,y) = e^{x^2+y^2}$, *y el dominio es* $D = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leqslant 1\}$.

Observemos que el dominio es un círculo de radio 1, *por lo que, en coordenadas polares, estos son los puntos con módulo menor que uno, y ángulos entre* 0 *y* 2π.

Por otro lado, x^2+y^2 *es el módulo al cuadrado* ρ^2, *de donde la función, luego del cambio de variables, queda* e^{ρ^2}.

Finalmente no debemos olvidar el determinante de la matriz Jacobiana del cambio de variables.

La integral es entonces:

$$\iint_D f(x,y)dxdy = \int_0^1 \int_0^{2\pi} e^{\rho^2} \rho \, d\theta d\rho.$$

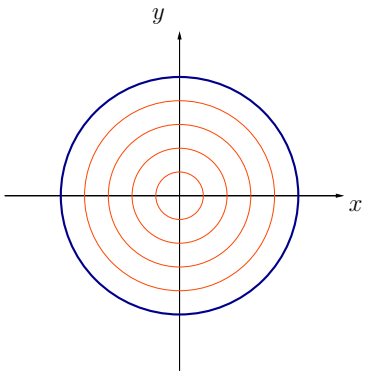
El resultado numérico es $\iint_D f(x,y)dxdy = \pi(e-1)$. *Queda como ejercicio terminar el cálculo.*

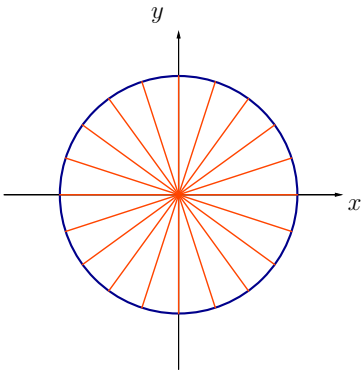
Volvamos a la formulación de la integral en coordenadas polares $\int_0^1 \int_0^{2\pi} e^{\rho^2} \rho \, d\theta d\rho$. *Observemos primero que, descripto en coordenadas polares, el dominio es el rectángulo* $[0,1] \times [0,2\pi]$, *y por lo tanto resulta igual de sencillo recorrerlo en un sentido o en otro. (estamos usando el Teorema de Fubini). Es decir, podemos plantear la integral de estas dos formas:*

$$\int_0^1 \int_0^{2\pi} e^{\rho^2} \rho \, d\theta d\rho = \int_0^{2\pi} \int_0^1 e^{\rho^2} \rho \, d\rho d\theta.$$

La interpretación geométrica en el dominio original de cada una de ellas, tiene diferencias conceptuales importantes.

Analicemos la primera de ellas, $\int_0^1 \int_0^{2\pi} e^{\rho^2}\rho$. Podemos leer esto de la siguiente forma: para cada valor de ρ entre 0 y 1, hacemos variar el ángulo θ entre 0 y 2π. Es decir, fijado un radio ρ, damos toda la vuelta, recorriendo una circunferencia. Como variamos el radio entre cero y uno, terminamos recorriendo todo el dominio.



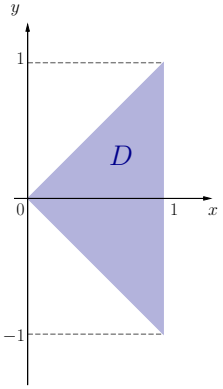


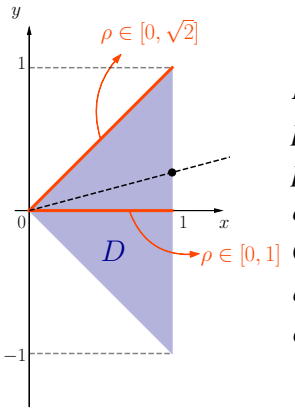
La segunda, $\int_0^{2\pi} \int_0^1 e^{\rho^2}\rho\, d\rho d\theta$, la podemos leer así: para cada ángulo θ, variamos el radio desde cero a uno. Es decir, fijado un ángulo θ, cubrimos un segmento de recta, radialmente. Como el ángulo varía entre 0 y 2π, terminamos barriendo con el segmento todo el dominio, en sentido anti-horario.

En este ejemplo era muy claro que el cambio a polares era conveniente: la función resulta más sencilla, y el dominio resulta mucho más sencillo. En otros casos no es tan claro, como en el ejemplo que veremos a continuación.

Ejemplo 7.15

Consideremos la integral de la función $f(x,y) = \frac{x^2}{x^2+y^2}$, en el dominio D de la figura.
Si realizamos un cambio de variable a polares, la función resulta $\frac{\rho^2\cos^2(\theta)}{\rho^2\cos^2(\theta)+\rho^2\sin^2(\theta)} = \cos^2(\theta)$.
Recorrer el dominio con coordenadas polares no parece inmediato. Recordemos que para cada ángulo θ_0 fijo, variando el módulo ρ, recorremos un segmento de recta.





En este caso, debemos recorrer ángulos entre $-\pi/4$ y $\pi/4$, pero para cada ángulo el módulo tiene límites distintos[7]*. Por ejemplo, para $\theta = 0$, el límite debe ser 1, pero para $\theta = \pi/4$ el módulo debe ir hasta $\sqrt{2}$.*
Consideremos entonces la recta $x = 1$, que es la arista del triángulo que oficia de "límite" para los valores del módulo ρ. Es decir, para cada ángulo θ, el módulo ρ debe ir desde cero hasta esta arista.

En coordenadas polares, esta recta $x = 1$ resulta $\rho\cos(\theta) = 1$. Es decir que podemos despejar el valor de ρ en función del ángulo θ, como $\rho = \frac{1}{\cos(\theta)}$.

Luego de el cambio de variables, la integral resulta entonces

$$\int_{-\pi/4}^{\pi/4}\int_{0}^{1/\cos(\theta)} \cos^2(\theta)\rho \, d\rho d\theta.$$

Queda como ejercicio terminar los cálculos, para corroborar que el resultado de la integral es $\pi/4$.

Ejercicio 7.16
Calcular la integral $\iint_D f$, donde $f(x,y) = \sqrt{x^2+y^2}$ y $D = \{(x,y) \in \mathbb{R}^2 : 0 \leqslant y, x^2+y^2 \geqslant 1, x^2+y^2-2x \leqslant 0\}$.

Integrales triples

De manera completamente análoga a lo visto en este capítulo, podemos definir integrales de funciones definidas en subconjuntos de $\mathbb{R}^3$.

Se generaliza la medida de Jordan (tomando pequeños cubos en vez de cuadrados), y luego las sumas inferiores, superiores, y la definición misma de integral, son idénticas.

Así, tenemos por ejemplo que el volumen del prisma $[0,A] \times [0,B] \times [0,C]$ es

$$\int_0^A \int_0^B \int_0^C 1 \, dzdydx = A \cdot B \cdot C.$$

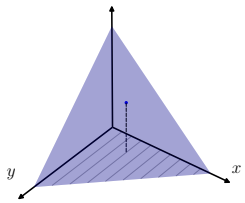
Veamos otro ejemplo un poco más interesante.

Ejemplo 7.17
Consideremos el conjunto $D = \{(x,y,z) \in \mathbb{R}^3 : x \geqslant 0, y \geqslant 0, z \geqslant 0, x+y+z \leqslant 1\}$, y la función $f(x,y,z) = xyz$.

[7] Observar que si el límite en ρ fuera igual para todos los ángulos, por ejemplo $\int_{-\pi/4}^{\pi/4}\int_0^1 \cdots d\rho d\theta$, estaríamos cubriendo una sección de círculo en vez del triángulo.

Para plantear la integral en este dominio, una opción es la siguente. Recorramos el triángulo del "piso", y luego para cada punto (x,y) de este plano, movemos z desde 0 hasta el "techo", que en este caso es el plano $x+y+z=1$.

Entonces, para recorrer el triángulo podemos hacer variar x entre cero y uno, y para cada valor de x, hacemos variar y entre 0 y $1-x$:



$$\int_0^1 \int_0^{1-x} \int_?^? f(x,y,z)\,dz\,dy\,dx.$$

Resta ahora definir los límites de integración para z. Entonces, para cada punto del triángulo, la altura varía entre 0 y $z=1-x-y$.

La integral triple entonces es

$$\int_0^1 \int_0^{1-x} \int_0^{1-x-y} xyz\,dz\,dy\,dx.$$

Como lo que resta son simples cálculos de primitivas y evaluaciones, se dejan como ejercicio.

Al igual que para integrales dobles, en muchas ocasiones resulta conveniente realizar un cambio de variables.

Veremos a continuación dos cambios de variable muy útiles. En el plano, el cambio a polares resultaba extremadamente conveniente en varias oportunidades, por lo que es natural preguntarse cuál es la generalización natural cuando trabajamos en $\mathbb{R}^3$.

Hay dos formas naturales de hacer esto:

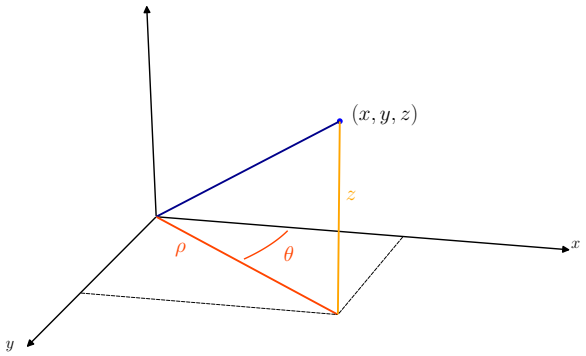
- Por un lado, podemos describir con coordenadas polares (un módulo ρ y un ángulo θ) cualquier punto del plano, y simplemente agregarle la altura z. Esto es lo que llamaremos **coordenadas cilíndricas**.
- Por otro lado, podemos pensar que en coordenadas polares lo que hacíamos era describir un punto indicando, por un lado, la dirección (para ello usábamos el ángulo θ), y por otro lado la distancia al origen (que indicábamos con ρ). En este sentido, podemos hacer algo similar en $\mathbb{R}^3$, para lo cual utilizaremos ρ para indicar la distancia al origen, pero ahora precisamos dos ángulos para indicar la dirección en el espacio. Llamaremos a este cambio, **coordenadas esféricas**.

7.1.2. Coordenadas Cilíndricas

Como adelantamos, en este cambio de variables utilizaremos dos nuevas variables (ρ y θ) que determinarán un módulo y un ángulo en el plano "del piso" $z = 0$, y luego utilizaremos la misma varible z para determinar la altura.

Es decir, el cambio de variable resulta:

$$\begin{aligned} x &= \rho\cos(\theta), \\ y &= \rho\sin(\theta), \\ z &= z. \end{aligned}$$



Otra forma de escribirlo es a través de la función g, que necesitaremos para calcular la Jacobiana. En este caso tenemos entonces $g : [0,\infty) \times [0,2\pi) \times \mathbb{R} \to \mathbb{R}^3$, definida como $g(\rho,\theta,z) = (\rho\cos(\theta), \rho\sin(\theta), z)$.

La matriz Jacobiana es

$$J_g(\rho,\theta,z) = \begin{pmatrix} \cos(\theta) & -\rho\sin(\theta) & 0 \\ \sin(\theta) & \rho\cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

de donde, desarrollando por la última fila o columna, tenemos que el determinante es el mismo que para el cambio a polares en el plano, es decir $|det(J_g(\rho,\theta,z))| = \rho$.

En este caso, los prismas en las coordenadas (ρ,θ,z) van a parar a cilindros en el espacio (x,y,z) (¿por qué?).

Veamos un ejemplo ilustrativo.

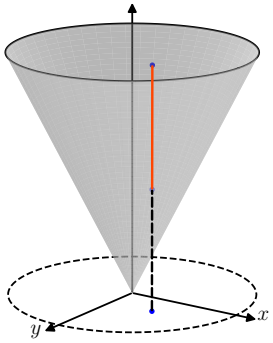
Ejemplo 7.18
Veamos cómo recorrer el dominio definido como $D = \{(x,y,z) \in \mathbb{R}^3 : \sqrt{x^2+y^2} \leqslant z \leqslant 1\}$, *utilizando coordenadas cilíndricas, de varias formas diferentes.*

Primero, observemos que el dominio es el interior del cono, como muestra la figura.

Para comenzar, con las variables ρ *y* θ *recorramos el círculo unidad ("en el piso", digamos). Esto ya lo hemos hecho con polares en* $\mathbb{R}^2$.

Hacemos variar entonces ρ entre cero y uno, y θ entre cero y 2π. Para cada punto del círculo, debemos recorrer con la altura z, de manera de cubrir el cono sólido. En este caso, para cada punto la altura debe variar entre la superficie dada por la ecuación del cono ($z = \sqrt{x^2+y^2}$) y el "techo", que es $z = 1$. Observemos que en coordenadas cilíndricas, la ecuación del cono es $z = \rho$.
Tenemos entonces:

$$\int_0^1 \int_0^{2\pi} \int_\rho^1 \cdots \rho \, dz \, d\theta \, d\rho$$

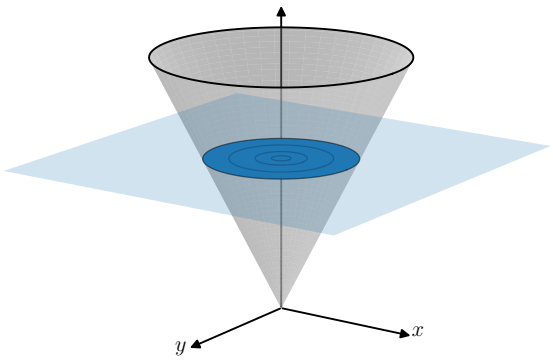


Otra forma de pensarlo es ir variando la altura z, y para cada uno de los planos horizontales resultantes, recorrer la sección correspondiente utilizando ρ y θ.

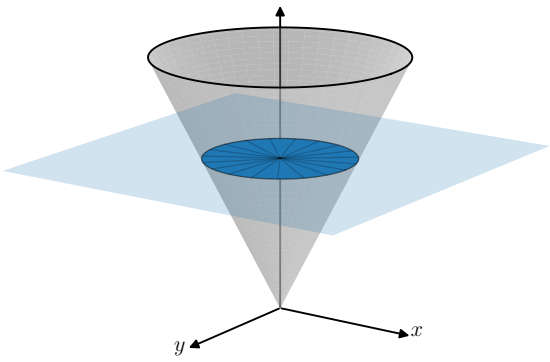
Observemos que para cada plano $z = cte$, la sección del dominio que tenemos que cubrir es un círculo de radio z (¿por qué?), y por lo tanto es muy sencillo recorrerlo con coordenadas polares. Específicamente, para cada altura z entre cero y uno, cubrimos el círculo mencionado haciendo variar θ de forma de dar una vuelta entera, y el módulo ρ desde cero hasta el radio correspondiente, que en este caso coincide con la altura z. Resulta entonces:

$$\int_0^1 \int_0^{2\pi} \int_0^z \cdots \rho \, d\rho \, d\theta \, dz$$

Observar que, tal como vimos en el ejemplo 7.14, para recorrer el círculo podemos intercambiar el orden de ρ y θ directamente, y esto corresponde a cubrir el dominio radialmente, o mediante circuferencias de distintos radios. La siguiente figura ilustra estas dos formas:



$$\int_0^1 \int_0^z \int_0^{2\pi} \cdots \rho \, d\theta \, d\rho \, dz$$



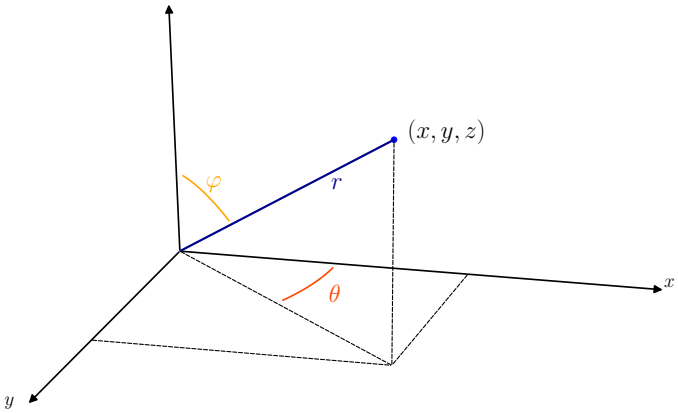
$$\int_0^1 \int_0^{2\pi} \int_0^z \cdots \rho \, d\rho \, d\theta \, dz$$

7.1.3. Coordenadas Esféricas

En coordenadas esféricas utilizaremos la variable r para indicar la distancia de un punto al origen, y dos ángulos para determinar la dirección.

Tomemos un punto (x,y,z) genérico. El cálculo de r es inmediato, y es $r = \sqrt{x^2+y^2+z^2}$. Ahora, existen varias formas de elegir un par de ángulos que determinen una dirección en el espacio[8].

Utilizaremos el ángulo φ para indicar el ángulo que forma el eje z con el vector (x,y,z) (ver figura). Estas dos direcciones, el eje z y el vector (x,y,z), forman un plano vertical. Utilizaremos el ángulo θ para indicar el ángulo que forma este plano con el eje x. Otra forma de verlo es proyectar el vector que forma el origen con el punto (x,y,z) en el plano $z=0$, y tomar el ángulo que forma esta proyección con el eje x.



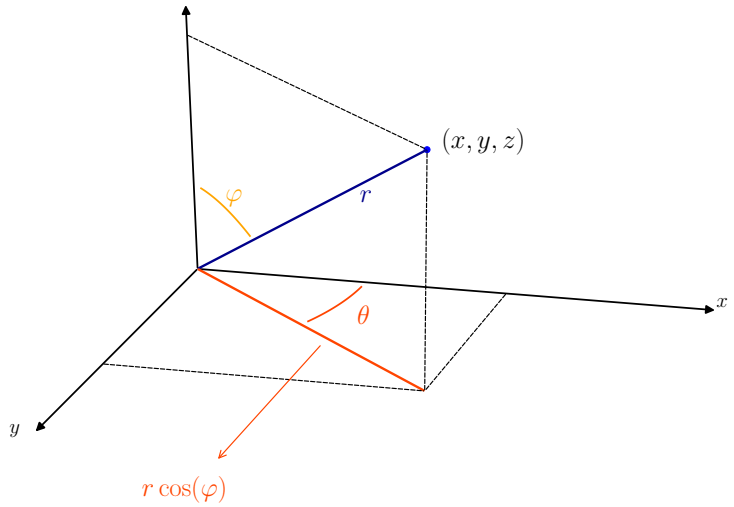
Veamos entonces cuál es la función de cambio de variables. Es decir, cómo se relacionan las viejas variables (x,y,z) con las nuevas variables (r,θ,φ).

La más sencilla es la variable z. Observemos para esto que en el plano que comentamos más arriba (que forma el eje z con el punto (x,y,z)), podemos proyectar el vector de módulo r que une el origen con el punto (x,y,z), sobre el eje z. Como este ángulo lo definimos como φ, tenemos que $z = r\cos(\varphi)$.

En el mismo plano, podemos proyectar el mismo vector sobre el "piso". Esta proyección, indicada en la figura, mide entonces $r\sin(\varphi)$. Luego, estamos en la misma situación que cuando definimos las coordendas polares, y debemos proyectar este vector de módulo $r\sin(\varphi)$ sobre los ejes, sabiendo que forma un ángulo θ con el eje x. Resulta entonces:

[8] Hay de hecho más de una convención para los ángulos en coordenadas esféricas

$$x = r\sin(\varphi)\cos(\theta),$$
$$y = r\sin(\varphi)\sin(\theta),$$
$$z = r\cos(\varphi).$$



Teniendo la función de cambio de variables, $g(r,\varphi,\theta) = (r\sin(\varphi)\cos(\theta), r\sin(\varphi)\sin(\theta), r\cos(\varphi))$, podemos calcular la matriz Jacobiana:

$$J_g(r,\varphi,\theta) = \begin{pmatrix} \sin(\varphi)\cos(\theta) & r\cos(\varphi)\cos(\theta) & -r\sin(\varphi)\sin(\theta) \\ \sin(\varphi)\sin(\theta) & r\cos(\varphi)\sin(\theta) & -r\sin(\varphi)\cos(\theta) \\ \cos(\varphi) & -r\sin(\varphi) & 0 \end{pmatrix}.$$

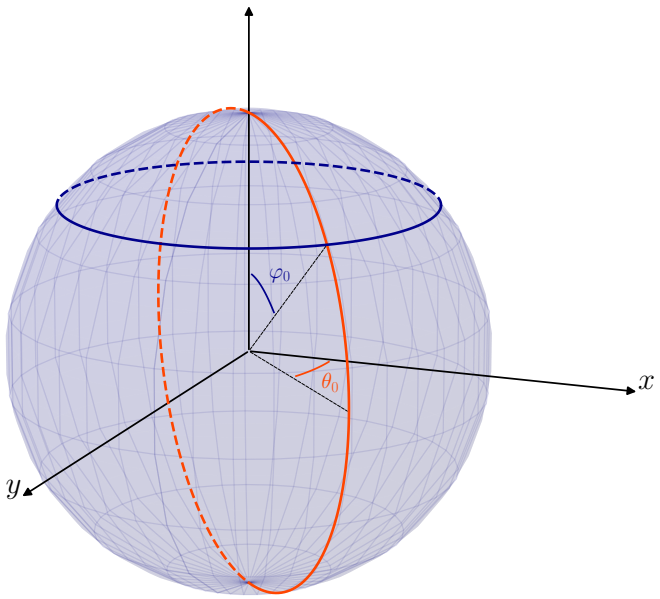
Ejercicio 7.19
Corroborar que $|\det(J_g(r,\varphi,\theta))| = r^2\sin(\varphi)$.

No vimos aún cuáles son los intervalos en los que deben variar las nuevas coordenadas (r,φ,θ), para poder describir cualquier punto del espacio R^3 (y sin repetir puntos, porque recordemos que una función de cambio de variables debe ser inyectiva[9]). Para ganar intuición sobre el cambio a esféricas y definir estos intervalos para las variables (r,φ,θ), estudiemos qué sucede en algunos casos particulares, suponiendo en principio que los ángulos pueden variar entre cero y 2π.

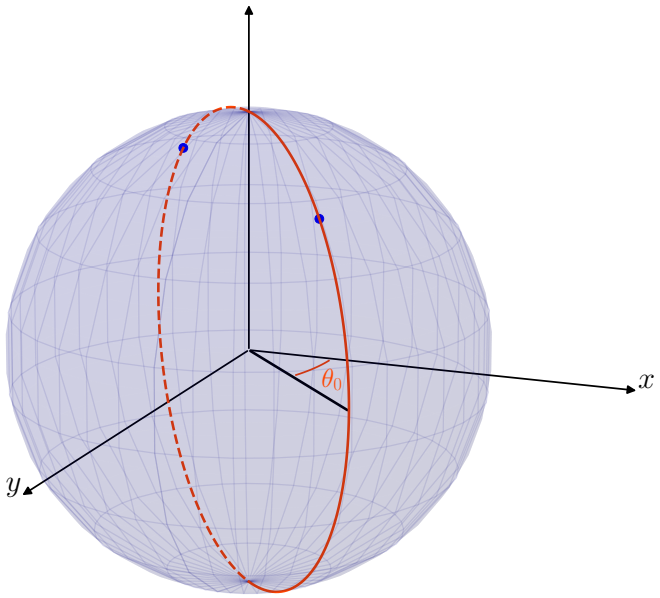
Para empezar, observemos que si fijamos los ángulos φ y θ (que determinan una dirección en el espacio), y dejamos variar r en $[0,+\infty)$, estamos recorriendo una semirrecta que comienza en el $(0,0,0)$. Este comportamiento es similar al que ya vimos en polares.

Fijemos ahora $r = 1$, que resulta en una esfera de radio 1, y estudiemos el comportamiento de los ángulos. Para empezar, fijemos también un ángulo $\theta = \theta_0$, y hagamos variar φ. El resultado es un "meridiano", como muestra la figura. Si, por el contrario, fijamos un ángulo φ_0, y hacemos variar θ, lo que obtenemos es la circunferencia en azul de la figura.

[9] En realidad esto se puede suavizar un poco. Observar por ejemplo lo que sucede en el cambio a polares. Si tomamos $\rho = 0$, no importa el ángulo θ, la imagen por g será el origen $(0,0)$. Es decir, todos los puntos de la forma $(0,\theta)$ van a parar al origen, por lo que claramente g no es inyectiva. De todas formas, como esto sucede solamente en un conjunto de medida cero, no genera problemas.



Ahora bien, volviendo al caso de θ_0 fijo, observemos los dos puntos que están marcados en la siguiente figura. En particular, supongamos que el punto que se encuentra en el primer octante (sobre la línea sólida) tiene módulo 1 y ángulos θ_0 y φ_0. Entonces, el punto "de atrás" (sobre la línea punteada), lo podemos representar de (por lo menos) dos formas. Por un lado, utilizando el mismo ángulo θ_0, pero ahora con un ángulo φ opuesto, es decir, $-\varphi_0$. Esto corresponde a moverse el mismo ángulo, pero en la dirección contraria, "hacia atrás". Por otro lado, podemos sumar media vuelta al ángulo θ_0, es decir, considerar un ángulo $\theta = \theta_0 + \pi$, y allí sí bajar un ángulo φ_0. En resumen, podemos representar al punto de atrás como $(1, -\varphi_0, \theta_0)$, o como $(1, \varphi_0, \theta_0 + \pi)$.



Esta observación es cierta para todo punto del espacio, y aquí sí tenemos problemas con la inyectividad entonces. Alcanza entonces considerar ángulos φ entre cero y π para poder describir cualquier punto del espacio, sin que existan conjuntos problemáticos de puntos con más de una representación[10].

Agregando los intervalos de variación, el cambio de variable es entonces:

$$\begin{aligned} x &= r\sin(\varphi)\cos(\theta), & r &\in [0,+\infty),\\ y &= r\sin(\varphi)\sin(\theta), & \varphi &\in [0,\pi],\\ z &= r\cos(\varphi). & \theta &\in [0,2\pi). \end{aligned}$$

Veamos un par de ejemplos.

Ejemplo 7.20
Consideremos la integral de la función $f(x,y,z) = z$, en el dominio definido como $D = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 \leqslant 1\}$.

El dominio es una esfera sólida, por lo que en coordenadas esféricas será sencillo de recorrer. Así como en coordenadas cilíndricas los prismas en las coordenadas nuevas van a parar a cilindros, en coordenadas esféricas los prismas van a parar a esferas. En efecto, la imagen de $[0,1]\times[0,\pi]\times[0,2\pi]$ es el dominio D.

La función luego del cambio de variables resulta $f(g(r,\varphi,\theta)) = r\cos(\varphi)$. Recordando que hay que multiplicar por el determinante de la matriz Jacobiana, la integral a resolver es

$$\iiint_D f = \int_0^1 \int_0^{2\pi} \int_0^{\pi} r\cos(\varphi) r^2 \sin(\varphi)\, d\varphi d\theta dr.$$

Observar que como los extremos de integración son constantes, podemos separar la integral como sigue:

$$\iiint_D f = \int_0^1 r^3 dr \cdot \int_0^{2\pi} d\theta \cdot \int_0^{\pi} \cos(\varphi)\sin(\varphi)\, d\varphi.$$

Realizando los cálculos, llegamos a que la primera integral (respecto de r) es $1/4$, la segunda integral (respecto de θ) es 2π, y la última es cero, por lo que el resultado final es $\iiint_D f = 0$.

En realidad ya podíamos prever este resultado, dada la simetría del dominio: la integral de z en la mitad superior de la esfera es opuesta a la integral de z en la mitad inferior.

[10] De todas formas, algunos (pocos) puntos todavía admiten más de una descripción en esféricas. ¿Cuáles?.

Ejemplo 7.21
Calculemos el volumen del conjunto $D = \{(x,y,z) \in \mathbb{R}^3 : \frac{1}{4} \leqslant \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 + \left(\frac{z}{c}\right)^2 \leqslant 1\}$*, donde a, b, y c son reales positivos.*

Para calcular el volumen, tenemos que calcular la integral de la función uno, en el dominio D.

Para deshacernos de las constantes, hagamos primero un cambio de variable lineal:

$$u = x/a,$$
$$v = y/b,$$
$$w = z/c.$$

Con este cambio, el nuevo dominio es $\tilde{D} = \{(u,v,w) \in \mathbb{R}^3 : \frac{1}{4} \leqslant u^2 + v^2 + w^2 \leqslant 1\}$*, que es más sencillo de interpretar. En efecto,* $\tilde{D}$ *es una esfera de radio uno, a la que le quitamos una esfera de radio un medio.*

Ahora, como hicimos un cambio de variable, debemos calcular el determinante de la matriz Jacobiana. En este caso, la función de cambio de variable es $g(u,v,w) = (ua, vb, wc)$*, por lo que la matriz Jacobiana es*

$$J_g(u,v,w) = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix},$$

de donde su determinante es abc.

Entonces, el volumen a calcular es $V = \iiint_{\tilde{D}} abc\, du\, dv\, dw$

Ahora sí, para realizar la integral en este nuevo dominio, hacemos el cambio a esféricas, y la integral resultante es ahora

$$V = abc \int_0^{2\pi} \int_0^{\pi} \int_{1/2}^{1} r^2 \sin(\varphi)\, dr\, d\varphi\, d\theta.$$

Es un ejercicio completar los cálculos para corroborar que el resultado es $V = abc\frac{4}{3}\frac{7}{8}\pi$.

Bibliografía

[1] T.M. Apostol. *Calculus, vol 1*. Reverte, 1975.

[2] R. Courant and F. John. *Introducción al cálculo y al análisis matemático*. Number v. 2. Limusa, 1996.

[3] IMERL. *Geometría y Álgebra Lineal 2 (libro rojo)*. IMERL, 2004.

[4] Israel Kleiner. Thinking the unthinkable: The story of complex numbers (with a moral). *Mathematics Teacher*, 81(7):583–92, 1988.

[5] Elon Lages Lima. *Espaços métricos*, volume 4. Instituto de Matemática Pura e Aplicada, CNPq Rio de Janeiro, 1983.

[6] Elon Lages Lima. *Análise Real, vol 1*. Number v. 1 in Coleção Matemática Universitária. Instituto de Matemática Pura e Aplicada, CNPq Rio de Janeiro, 2016.

[7] Elon Lages Lima. *Análise Real, vol 2*. Number v. 2 in Coleção Matemática Universitária. Instituto de Matemática Pura e Aplicada, CNPq Rio de Janeiro, 2017.

[8] Alexandre Miquel. *Notas de Cálculo Diferencial e Integral en una variable*. IMERL, 2017.

[9] M. Spivak. *Calculus On Manifolds: A Modern Approach To Classical Theorems Of Advanced Calculus*. Mathematics monograph series. Avalon Publishing, 1971.

[10] M. Spivak. *Calculus*. Calculus. Cambridge University Press, 2006.